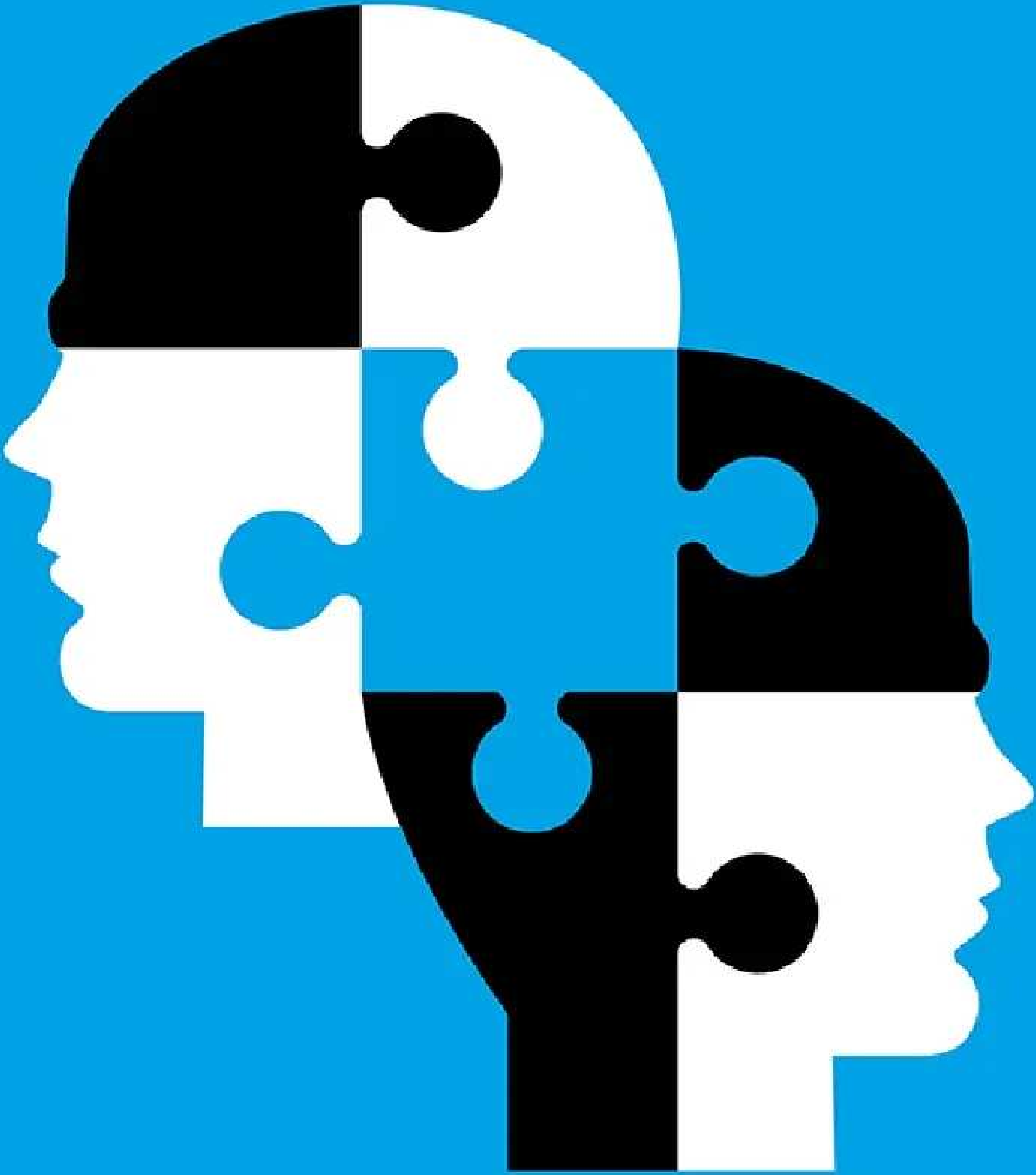

ELEVENTH EDITION

How to Think Straight About Psychology

Keith E. Stanovich

How to Think Straight About Psychology

ELEVENTH EDITION

Keith E. Stanovich
University of Toronto

Pearson

330 Hudson Street, NY NY 10013

Portfolio Manager: Tanimaa Mehra
Content Producer: Kani Kapoor
Portfolio Manager Assistant: Anna Austin
Product Marketer: Jessica Quaza
Content Producer Manager: Jennifer Sargunar
Art/Designer: Integra Software Services Pvt. Ltd.
Full-Service Project Manager: Integra Software Services Pvt. Ltd.
Compositor: Integra Software Services Pvt. Ltd.
Printer/Binder: LSC Communications, Inc.
Cover Printer: Phoenix Color/Hagerstown
Cover Design: Lumina Datamatics, Inc.
Cover Art: Shutterstock

Library of Congress Cataloging-in-Publication Data
Names: Stanovich, Keith E., 1950- author.
Title: How to think straight about psychology / Keith E. Stanovich, University of Toronto.
Description: Eleventh edition. | Hoboken, New Jersey: Pearson Education, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017037900| ISBN 9780134478623 | ISBN 0134478622
Subjects: LCSH: Psychology—Research—Methodology. | Mass media—Psychological aspects. | Mass media—Objectivity.
Classification: LCC BF76.5 .S68 2018 | DDC 150.72—dc23 LC record available at https://lccn.loc.gov/2017037900

1 17

Books a la Carte
ISBN-10: 0-13-447862-2
ISBN-13: 978-0-13-447862-3

Pearson

To Paula, who taught me how to think straight about life

Contents

Preface

There exists a body of knowledge that is unknown to most people. This information concerns human behavior and consciousness in their various forms. It can be used to explain, predict, and control human actions. Those who have access to this knowledge use it to gain an understanding of other human beings. They have a more complete and accurate conception of what determines the behavior and thoughts of other individuals than do those who do not have this knowledge.

Surprisingly enough, this unknown body of knowledge is the discipline of psychology.

What can I possibly mean when I say that the discipline of psychology is unknown? Surely, you may be thinking, this statement was not meant to be taken literally. Bookstores contain large sections full of titles dealing with psychology. Television and radio talk shows regularly feature psychological topics. Magazine articles and websites quote people called psychologists talking about a variety of topics. Yet, despite all of this, there is an important sense in which the *field* of psychology is unknown.

The transfer of "psychological" knowledge that is taking place via the media is largely an illusion. Few people are aware that the majority of the books they see in the psychology sections of many bookstores are written by individuals with absolutely no standing in the psychological community. Few are aware that many of the most visible psychological "experts" have contributed no information to the fund of knowledge in the discipline of psychology.

The flurry of media attention paid to "psychological" topics has done more than simply present inaccurate information. It has also obscured the very real and growing knowledge base in the field of psychology. The general public is unsure about what is and is not psychology and is unable to independently evaluate claims about human behavior. Adding to this problem is the fact that many people have a vested interest in a public that is either without evaluative skills or that believes there is no way to evaluate psychological claims. The latter view, sometimes called the "anything goes" attitude, is one of the fallacies discussed in this book, and it is particularly costly to the public. Many pseudosciences are multimillion-dollar industries that depend on the lack of public awareness that claims about human behavior can be tested. The general public is also unaware that many of the claims made by these pseudosciences (e.g., astrology, psychic surgery, speed reading, biorhythms, therapeutic touch, subliminal self-help tapes, facilitated communication, indigo children, psychic detectives) have been tested and proved false. The existence of the pseudoscience industry, which is discussed in this book, increases the media's tendency toward sensationalistic reporting of science. This tendency is worse in psychology than in other sciences, and understanding the reasons why this is so is an important part of learning how to think straight about psychology.

This book, then, is directed not at potential researchers in psychology but at a much larger group: the consumers of psychological information. The target audience is the beginning psychology student and the general reader who have encountered information on psychological issues in the general media and have wondered how to go about evaluating its validity.

This book is not a standard introductory psychology text. It does not outline a list of facts that psychological research has uncovered. Indeed, telling everyone to take an introductory psychology course at a university is probably not the ultimate solution to the inaccurate portrayal of psychology in the media. There are many laypeople with a legitimate interest in psychology who do not have the time, money, or access to a university to pursue formal study. More importantly, as a teacher of university-level psychology courses, I am forced to admit that my colleagues and I often fail to give our beginning students a true understanding of the science of psychology. The reason is that lower-level courses often do not teach the critical analytical skills that are the focus of this book. As instructors, we often become obsessed with "content"—with "covering material." Every time we stray a little from the syllabus to discuss issues such as psychology in the media, we feel a little guilty and begin to worry that we may not cover all the topics before the end of the term.

Consider the average introductory psychology textbook. Many now contain between 600 and 800 multicolumned pages and reference literally hundreds of studies in the published literature. Of course, there is nothing wrong with such books containing so much material. It simply reflects the increasing knowledge base in psychology. There are, however, some unfortunate side effects. Instructors are often so busy trying to cram their students full of dozens of theories, facts, and experiments that they fail to deal with some of the fundamental questions and misconceptions that students bring with them to the study of psychology. Rather than dealing directly with these misconceptions, the instructors (and the introductory textbook authors) often hope that if students are exposed to enough of the empirical content of psychology, they will simply *induce* the answers to their questions. All too often this hope is frustrated. In

a final review session—or in office hours at the end of the term—instructors are often shocked and discouraged by questions and comments that might have been expected on the first day of the course but not after 14 weeks: "But psychology experiments aren't real life; what can they tell us?"; "Psychology can't be a *real* science like chemistry, can it?"; "But I heard a therapist on TV say the opposite of what our textbook said"; "I think this theory is stupid—my brother behaves just the opposite of what it says"; "Psychology is nothing more than common sense, isn't it?"; "Everyone knows what anxiety is—why bother defining it?" For many students, such questions are not implicitly answered merely by a consideration of the content of psychology. In this book, I deal explicitly with the confusions that underlie questions and comments such as these.

Unfortunately, research has shown that the average introductory psychology course does surprisingly little to correct students' misconceptions about the discipline (Kowalski & Taylor, 2009; Lilienfeld, 2014; Taylor & Kowalski, 2004). This unfortunate fact provides the rationale for this book. Psychology students need explicit instruction in the critical thinking skills that will transform them into independent evaluators of psychological information.

Years after students have forgotten the content of an introductory psychology course, they will still use the fundamental principles covered in this book to evaluate psychological claims. Long after Erikson's stages of development have been forgotten, students will be using the thinking tools introduced in this text to evaluate new psychological information encountered in the media. Once acquired, these skills will serve as lifelong tools that will aid in the evaluation of knowledge claims. For example, these skills provide some criteria for assessing the reliability of "expert" opinion. Because the need to rely on expert opinion can never be eliminated in a complex society, the evaluation of an expert's credibility becomes essential to knowledge acquisition.

Many psychologists are pessimistic about any effort to stem the tide of misinformation about their discipline. Although this pessimism is, unfortunately, often justified, this "consumer's guide" to psychology was motivated by the idea that psychologists must not let this problem become a self-fulfilling prophecy.

Although I have welcomed the opportunity to prepare several editions of *How to Think Straight About Psychology*, it is unfortunately true that the reasons for the book's existence are just as applicable today as they were when I wrote the first edition. Students in introductory psychology courses enter with as many misconceptions as they ever did, and they think that unaided common sense is all they need to understand human behavior, or worse, they turn to pseudosciences. Thus, the goals of all subsequent editions have remained the same: to present a short introduction to the critical thinking skills that will help students to better understand the subject matter of psychology.

New to the Edition

The eleventh edition of *How to Think Straight About Psychology* has no major structural revisions because a chapter reorganization occurred in a previous edition. The content and order of the chapters remain the same. At the request of reviewers and users, this edition remains at the same length as the tenth edition. Readers and users have not wanted the book to lengthen and, indeed, it has not. I have continued to update and revise the examples that are used in the book (while keeping those that are reader favorites). Some dated examples have been replaced with more contemporary studies and issues. I have made a major effort to use contemporary citations that are relevant to the various concepts and experimental effects that are mentioned. A large number of new citations appear in this edition (290 new citations, to be exact!), so that the reader continues to have up-to-date references on all of the examples and concepts.

New examples, discussions, and sections have been added. A sampling of these new additions include the following issues and discussions: cell phone use while driving; the use of psychology in child custody disputes; pseudoscience in clinical psychology; the efficacy of crisis counselling after traumatic events; the causes of people making bad investment decisions; the "reading wars" in education; the effects of violent video games; facilitating communication in autism; conducting experiments over the Internet; the left-brain/right brain fallacy; health outcomes of alcohol consumption; distraction from electronic dashboard devices; coverage of the replication crisis in psychology; a new emphasis on the evils of vanity publishing; an additional section on the relation between lab and field results in psychology; a discussion of the Amazon Mechanical Turk; a discussion of how vivid presentations of results from neuroscience can skew conclusions; a discussion of the fallacies surrounding the mistaken idea of multitasking; a new discussion of Walter Mischel's famous marshmallow studies and how they exemplify going from basic to applied research; a discussion of the danger of the phrase "new study shows" in the media; many additional examples of the use of meta-analytic studies in psychology (including marriage longevity, brain training, predictors of job performance, and suicide prevention); a discussion of how the media suggest that science is non-cumulative in their reporting of research on autism and reading disability and ADHD.

The goal of the book remains what it always was—to present a short introduction to the critical thinking skills that will help the student to better understand the subject matter of psychology. During the past decade and a half there has been an increased emphasis on the teaching of critical thinking in universities (Arum & Roksa, 2011; Sternberg, Roediger, & Halpern, 2006). Indeed, some state university systems have instituted curricular changes mandating an

emphasis on critical thinking skills. At the same time, however, other educational scholars were arguing that critical thinking skills should not be isolated from specific factual content. *How to Think Straight About Psychology* combines these two trends. It is designed to provide the instructor with the opportunity to teach critical thinking within the rich content of modern psychology.

Readers are encouraged to send me comments at: keith.stanovich@utoronto.ca.

Acknowledgments

Many of the individuals I have acknowledged in earlier editions continue to contribute ideas for the book. However, I must single out Richard West of James Madison University, who has been a most valuable continuing contributor to the book's evolution. A humane scholar and a true friend, his intellectual and emotional support is much appreciated.

Several other scholars have provided valuable feedback on this and earlier editions. These include Wayne Bartz, American River College; Christopher Bauer, University of New Hampshire; Ludy Benjamin, Texas A&M University; Angela M. Birkhead-Flight, University of Cincinnati; Virginia Blankenship, University of Northern Arizona; Edward C. Chang, Northern Kentucky University; Michael Choban, West Virginia Wesleyan University; James Clark, University of Winnipeg; Jim Coan, University of Arizona; Ellen Cole, Alaska Pacific University; Ellen Cotter, Georgia Southwestern State University; Anne Cunningham, University of California, Berkeley; Ian Deary, University of Edinburgh; Julie Deisinger, Saint Xavier University; David DiBattista, Brock University; Wallace Dixon, Heidelberg College; Mark Fineman, Southern Connecticut State University; Herbert Fink, SUNY–Brockport; Heinz Fischer, Long Beach City College; Ronald Gandelman, Rutgers University; Michael Gasser, University of Northern Iowa; Traci A. Giuliano, Southwestern University; William Graziano, Purdue University; Nancy J. Gussett, Baldwin-Wallace College; Gordon Hammerle, Adrian College; Randy Hansen, Oakland University; William L. Hathaway, Regent University; George Heise, Indiana University; Albert Heldt, Grand Rapids Junior College; Dori Henderson, Metropolitan State University; George Howard, University of Notre Dame; Barry Kendall, Ontario, Canada; Bernie Koenig, Fanshawe College; Victor Koop, Goshen College; Andy Kwong, University of New South Wales; P. A. Lamal, University of North Carolina, Charlotte; Stephen Louisell, Kalamazoo Community College; Gwen Lupfer-Johnson, University of Alaska, Anchorage; Margaret Matlin, State University of New York-Geneseo; Douglas Mook, University of Virginia; Timothy Moore, York University; Edward Morris, University of Kansas; Joseph E. Morrow, California State University at Sacramento; Michael O'Boyle, Iowa State University; Blaine Peden, University of Wisconsin, Eau Claire; John F. Pfister, Dartmouth College; Sam Rakover, University of Haifa; Richard Redding, Hahneman University; Michael Ross, University of Waterloo; John Ruscio, Elizabethtown College; Walter Sa, Grand Valley State University; Allen Salo, University of Maine at Presque Isle; Frank Schieber, University of South Dakota; Mike Seiler, Oakland University; Jillene Grover Seiver, Bellevue College; Marjorie Semonick, University of Minnesota; David Share, University of Haifa; Jeffrey Sherman, Northwestern University; Linda Siegel, University of British Columbia; Norman Silverman, University of Illinois, Chicago; Frank Smoll, University of Washington; Paul Solomon, Williams College; Mike Stadler, University of Missouri; Maggie Toplak, York University; Larry Vandervert, Spokane Falls Community College; John Vokey, University of Lethbridge; Carol Wade, College of Marin; Marty Wall, University of Toronto; Barbara Wanchisen, Baldwin-Wallace College; Toni G. Wegner, University of Virginia; Edward Wisniewski, Northwestern University; Murray S. Work, California State University at Sacramento; and Edward Zuckerman, Guilford Press.

Special thanks to Gerald L. Peterson from Saginaw Valley State University for some very insightful comments.

The insights from many discussions about teaching methodology with Ted Landau, Larry Lilliston, and Dean Purcell, all of Oakland University, were incorporated into the book. Reviewers of recent editions who were particularly helpful include Michael Choban, West Virginia Wesleyan University; David DiBattista, Brock University; Steven Isonio, Golden West College; John Ruscio, Elizabethtown College; Allen Salo, University of Maine at Presque Isle; Cindy Sifonis, Oakland University; Michael Tagler, Nebraska Wesleyan University; and Chris Ward, Stonehill College.

My editor at Pearson, Priya Christopher, has provided guidance, enthusiasm, and support for the book. The production process benefitted from the support of: Erin Mitchell, Carly Czech, Anita Castro, Courtney Welsh, Allison Campbell, Anju Baskar, and Joel Morgan Kearney. Robyn Macpherson is thanked for her diligent library and reference assistance on several editions prior to this one.

Finally, I wish to thank Paula J. Stanovich for more than just the emotional support that is routinely alluded to in acknowledgments. Her concern for all human beings, particularly those less fortunate, is an inspiration to all who know her. A view we both share is that all human beings should have the opportunity to utilize their full potential. This book attests to the fact that I have had such an opportunity. Paula works to speed the day when this opportunity will be fully extended to all individuals with disabilities.

Chapter 1

Psychology Is Alive and Well (and Doing Fine Among the Sciences)

Learning Objectives

1.1 Explain why Freud's methods are unrepresentative of modern psychology

1.2 Describe the implications of diversity in the field of psychology

1.3 Differentiate psychology from other disciplines that deal with human behavior

1.4 Describe the three features that define science

1.5 Distinguish between psychology and folk wisdom.

1.6 Explain the reasons for the hostility directed towards psychology as a discipline

The Freud Problem

Stop 100 people on the street and ask them to name a psychologist, either living or dead. Record the responses. Of course, Dr. Phil and other "media psychologists" would certainly be named. If we leave out the media and pop psychologists, however, and consider only those who have had an impact on psychology as a discipline, there would be no question about the outcome of this informal survey. Sigmund Freud would be the winner hands down. B. F. Skinner would finish a distant second (Roediger, 2016; Sternberg, 2016). No other psychologist would get enough recognition even to bother about. Thus, Freud, along with the pop psychology presented in the media, largely defines psychology in the public mind.

The notoriety of Freud has greatly affected the general public's views of psychology and has contributed to many misunderstandings. For example, many introductory psychology students are surprised to learn that if all the members of the American Psychological Association (APA) who were concerned with Freudian psychoanalysis were collected, they would make up less than 5 percent of the membership (Engel, 2008). In another major psychological association, the Association for Psychological Science, they would be even less common. One popular introductory psychology textbook (Wade & Tavris, 2008) is over 700 pages long, yet contains only 15 pages on which either Freud or psychoanalysis is mentioned—and these 15 pages often contain

criticism ("most Freudian concepts were, and still are, rejected by most empirically oriented psychologists," p. 19). Developmental psychologist Alison Gopnik (2014) calls Freudian theory a zombie idea that has haunted English departments of universities long after it had nearly disappeared from psychology.

In short, modern psychology is not obsessed with the ideas of Sigmund Freud, nor is it largely defined by them. Freud's work is an extremely small part of the varied set of issues, data, and theories that concern modern psychologists. This larger body of research and theory encompasses the work of 5 Nobel Prize winners (David Hubel, Daniel Kahneman, Herbert Simon, Roger Sperry, and Torsten Wiesel) and 17 winners of the National Medal of Science (Lowman & Benjamin, 2012), all of whom are virtually unknown to the public.

It is bad enough that Freud's importance to modern psychology is vastly exaggerated. What makes the situation worse is that Freud's methods of investigation are completely unrepresentative of how modern psychologists conduct their research. In fact, Freud's methods give an utterly *misleading* impression of psychological research. For example, Freud did not use controlled experimentation, which, as we shall see in Chapter 6, is the most potent weapon in the modern psychologist's arsenal of methods. Freud thought that case studies could establish the truth or falsity of theories. We shall see in Chapter 4 why this idea is mistaken. As one historian of psychotherapy has noted, "If Freud himself was a scientist, it was a strange science he was promulgating. . . . Psychoanalysis contained theories and hypotheses, but it lacked a method of empirical observation" (Engel, 2008, p. 17).

Finally, a critical problem with Freud's work concerns the connection between theory and behavioral data. As we shall see in Chapter 2, for a theory to be considered scientific, the link between the theory and behavioral data must meet some minimal requirements. Freud's theories do not meet these criteria (Boudry & Buekens, 2011; Dufresne, 2007; Engel, 2008). To make a long story short, Freud built an elaborate theory on a database (case studies and introspection) that was not substantial enough to support it. Freud concentrated on building complicated theoretical structures, but he did not, as modern psychologists do, ensure that they would rest on a database of reliable, replicable behavioral relationships. In summary, familiarity with Freud's style of work can be a significant impediment to the understanding of modern psychology.

In this chapter, we shall deal with the Freud problem in two ways. First, when we illustrate the diversity of modern psychology, the rather minor position occupied by Freud will become clear. Second, we shall discuss what features are common to psychological investigations across a wide variety of domains (features missing from Freud's work). We will see that there is one unifying characteristic of modern psychology: the quest to understand behavior by using the methods of science.

The Diversity of Modern Psychology

There is, in fact, a great diversity of content in modern psychology. A textbook once referred to psychology as "a loosely federated intellectual empire that stretches from the domains of the biological sciences on one border to those of the social sciences on the other" (p. 774, Gleitman, 1981). Understanding that psychology is composed of an incredibly wide and diverse set of investigations is critical to an appreciation of the nature of the discipline. Simply presenting some of the concrete indications of this diversity will illustrate the point. The APA has 54 different divisions, each representing either a particular area of research or a particular area of practice (see Table 1.1). From the table, you can see the range of subjects studied by psychologists, the range of settings involved, and the different aspects of behavior studied. The other large organization of psychologists—the Association for Psychological Science—is just as diverse. Actually, Table 1.1 understates the diversity within the field of psychology because it gives the impression that each division is a specific specialty area. In fact, each of the 54 divisions listed in the table is a broad area of study that contains a wide variety of subdivisions! In short, it is difficult to exaggerate the diversity of the topics that fall within the field of psychology.

Table 1.1 Divisions of the American Psychological Association

1. General Psychology	30. Psychological Hypnosis
2. Teaching of Psychology	31. State Psychological Association Affairs
3. Experimental Psychology	32. Humanistic Psychology
5. Evaluation, Measurement, and Statistics	33. Intellectual and Developmental Disabilities
6. Behavioral Neuroscience and Comparative Psychology	34. Population and Environmental Psychology
7. Developmental Psychology	35. Psychology of Women
8. Personality and Social Psychology	36. Psychology of Religion
9. Psychological Study of Social Issues	37. Child and Family Policy and Practice
10. Psychology of Aesthetics, Creativity, and the Arts	38. Health Psychology
12. Clinical Psychology	39. Psychoanalysis
13. Consulting Psychology	40. Clinical Neuropsychology
14. Industrial and Organizational Psychology	41. Psychology and Law
15. Educational Psychology	42. Psychologists in Independent Practice
16. School Psychology	43. Family Psychology
17. Counseling Psychology	44. Psychological Study of Lesbian, Gay, and Bisexual Issues
18. Psychologists in Public Service	45. Psychological Study of Ethnic Minority Issues
19. Military Psychology	46. Media Psychology
20. Adult Development and Aging	47. Exercise and Sport Psychology
21. Applied Experimental and Engineering Psychology	48. Peace Psychology
22. Rehabilitation Psychology	49. Group Psychology and Group Psychotherapy
23. Consumer Psychology	50. Addictions
24. Theoretical and Philosophical Psychology	51. Psychological Study of Men and Masculinity
25. Behavior Analysis	52. International Psychology
26. History of Psychology	53. Clinical Child and Adolescent Psychology
27. Community Psychology	54. Pediatric Psychology
28. Psychopharmacology and Substance Abuse	55. Pharmacotherapy
29. Psychotherapy	56. Trauma Psychology

NOTE: There is no Division 4 or 11.

Implications of Diversity

Many people come to the study of psychology hoping to learn the one grand psychological theory that unifies and explains all aspects of human behavior. Such hopes are often disappointed, because psychology contains not one grand theory, but many different theories, each covering a limited aspect of behavior. The diversity of psychology guarantees that the task of theoretical unification will be immensely difficult. Indeed, many psychologists would argue that such a unification is impossible. Others, however, are searching for greater unification within the field (Brewer, 2013; Schwartz et al., 2016; Simonton, 2015). For example, the coherence of psychology as a discipline has increased over the last three decades due to the theoretical efforts of evolutionary psychologists. These researchers have tried to bring unification to our conceptualization of human psychological processes by viewing them as mechanisms serving critical evolutionary functions such as kinship recognition, mate selection, cooperation, social exchange, and child rearing (Buss, 2011; Cartwright, 2016; Geary, 2005, 2009). Likewise, Cacioppo (2007) points to subfields such as social cognitive neuroscience as tying together numerous specialty areas within psychology—in this case, cognitive psychology, social psychology, and neuropsychology.

No matter what their position on the issue of the coherence of the subject matter of psychology, all psychologists agree that theoretical unification will be an extremely difficult task. The lack of theoretical integration leads some critics of psychology to denigrate the scientific progress that psychology has made. Such criticism often arises from the mistaken notion that all true sciences must have a grand, unifying theory. It is a mistaken notion because many other sciences also lack a unifying conceptualization.

Some scholars have argued that the term "*psychology*" implies a coherence of subject matter that is not characteristic of the discipline. As a result, a number of leading university departments in the United States have been changing their names to Department of Psychological Sciences (Jaffe, 2011; Klatzky, 2012). The term "sciences" conveys two of the important messages of this chapter. That the term is plural signals the point about the diversity of content in the discipline that we have been discussing. The term "sciences" also signals where to look for the unity in the discipline of psychology—not to its content, but instead to its *methods.* Here is where we actually find more unity of purpose among psychologists.

Unity in Science

Simply to say that psychology is concerned with human behavior does not distinguish it from other disciplines. Many other professional groups and disciplines—including economists, novelists, the law, sociology, history, political science, and anthropology—are, in part, concerned with human behavior. Psychology is not unique in this respect.

Practical applications do not establish any uniqueness for the discipline of psychology either. For example, many university students decide to major in psychology because they have the laudable goal of wanting to "help people." But helping people is an applied part of an incredibly large number of fields, including social work, education, nursing, occupational therapy, physical therapy, police science, human resources, and speech therapy. Similarly, the goal of training applied specialists to help people by counseling them does not demand that we have a discipline called psychology. Helping people by counseling them is an established part of many other fields, including education, social work, police work, nursing, pastoral work, occupational therapy, and many others.

It is easy to argue that there are really *only* two things that justify psychology as an independent discipline. The first is that psychology studies the full range of human and nonhuman behavior with the techniques of science. The second is that the applications that derive from this knowledge are *scientifically* based. Were this not true, there would be no reason for psychology to exist.

Psychology is different from other behavioral fields in that it attempts to give the public two guarantees. One is that the conclusions about behavior that it produces derive from scientific evidence. The second is that practical applications of psychology have been derived from and tested by scientific methods. Does psychology ever fall short of these goals? Yes, quite often (Duarte et al., 2015; Ferguson, 2015; Lilienfeld, 2012). This book is about how we might better attain them. I will return in Chapter 12 to the issue of psychologists themselves undermining their own legitimacy by not meeting appropriate scientific standards. But, in principle, these are the standards that justify psychology as an independent field. If psychology ever decides that these goals are not worth pursuing—that it does not wish to adhere to scientific standards—then it might as well fold its tent and let its various concerns devolve to other disciplines because it would be a totally redundant field of intellectual inquiry.

Clearly, then, the first and most important step that anyone must take in understanding psychology is to realize that its defining feature is that it is the *data-based scientific study of behavior*. Comprehending all of the implications of this fact will occupy us for the rest of this book, because it is the primary way that we develop the ability to think straight about psychology. Conversely, the primary way that people get confused in their thinking about psychology is that they often fail to realize that it is a scientific discipline. For example, it is quite common to hear people outside the discipline voice the opinion that psychology is not a science. Why is this a common occurrence?

Attempts to convince the public that psychology cannot be a science stem from a variety of sources. For example, there currently exist many industries surrounding

pseudoscientific belief systems that have a vested interest in convincing the public that anything goes in psychology and that there are no rational criteria for evaluating psychological claims. This is the perfect atmosphere in which to market claims like: "Lose weight through hypnosis," "Develop your hidden psychic powers," and "Learn French while you sleep," along with the many other parts of the multibillion-dollar self-help industry that either are not based on scientific evidence or, in many cases, are actually contradicted by much available evidence.

Another source of resistance to scientific psychology stems from the tendency to oppose the expansion of science into areas where unquestioned authorities and "common sense" have long reigned. History provides many examples of initial public resistance to the use of science rather than philosophical speculation, theological edict, or folk wisdom to explain the natural world. Each science has gone through a phase of resistance to its development. Learned contemporaries of Galileo refused to look into his new telescope because the existence of the moons of Jupiter would have violated their philosophical and theological beliefs. For centuries, the understanding of human anatomy progressed only haltingly because of prohibitions on the dissection of human cadavers. Charles Darwin was repeatedly denounced. Paul Broca's Society of Anthropology was opposed in France in the nineteenth century because knowledge about human beings was thought to be subversive to the state.

Each scientific step to greater knowledge about human beings has evoked opposition. This opposition eventually dissipated, however, when people came to realize that science does not destroy the meaning of our lives but enhances it. Who now believes that astronomy's mapping of the galaxies and its intricate theories about the composition of distant stars destroy our wonder at the universe? Who would substitute the health care available in their community for that available before human cadavers were routinely dissected? An empirical attitude toward the stars or the human body has not diminished humanity. More recently, Darwin's evolutionary synthesis laid the foundation for startling advances in biology. Nevertheless, as we get closer to the nature of human beings and their origins, vestiges of opposition remain. Many people remain uncomfortable with the implications of evolutionary theory (Dennett, 1995; Stanovich, 2004). If evolutionary biology, with its long and impressive record of scientific achievements, still engenders public opposition, is it any wonder that psychology, the most recent discipline to bring long-held beliefs about human beings under scientific scrutiny, currently provokes people to deny its validity?

What, Then, Is Science?

In order to understand what psychology is, we must understand what science is. We can begin by dealing with what science is not. First, science is not defined by subject matter. Any aspect of the universe is fair game for the development of a scientific discipline, including all aspects of human behavior. We cannot divide the universe into "scientific" and "nonscientific" topics. Although strong forces throughout history have tried to place human beings outside the sphere of scientific investigation, they have been unsuccessful, as we shall see. The reactions against psychology as a scientific discipline probably represent the modern remnants of this ancient struggle.

Science is also not defined by the presence of instruments and experimental apparatus. It is not the test tube, the computer, the electronic equipment, or the investigator's white coat that defines science. These are the trappings of science but are not its defining features. Science is, rather, a way of thinking about and observing the universe that leads to a deep understanding of its workings.

In the remainder of this chapter, we shall discuss three important and interrelated features that define science: (1) the use of systematic empiricism; (2) the production of public knowledge; and (3) the examination of solvable problems. Although we shall examine

each feature separately, remember that the three connect to form a coherent general structure. (For a more detailed discussion of the general characteristics of a science, see the works of Bronowski, Medawar, and Popper listed in the references section of this book.)

Systematic Empiricism

If you look up the word empiricism in any dictionary, you will find that it means "the practice of relying on observation." Scientists find out about the world by examining it. The fact that this point may seem obvious to you is an indication of the spread of the scientific attitude in the past couple of centuries. In the past, it has not always seemed so obvious. Recall the example of Galileo. With his primitive telescope, Galileo claimed to have seen moons around the planet Jupiter at a time when it was thought by learned people that there were only seven "heavenly bodies" (five planets, the sun, and the moon). This was at a time when it was thought that knowledge was best obtained through pure thought or through appeal to authority. Some contemporary scholars refused to look into Galileo's telescope. Others said the telescope was designed to trick. Still others said that it worked on Earth but not in the sky (Shermer, 2011). Another scholar, Francesco Sizi, attempted to refute Galileo not with observations, but with the following argument:

> There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which and many other similar phenomena of nature such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven.... Besides, the Jews and other ancient nations, as well as modern Europeans, have adopted the division of the week into seven days, and have named them from the seven planets; now if we increase the number of planets, this whole system falls to the ground.... Moreover, the satellites are invisible to the naked eye and therefore can have no influence on the earth and therefore would be useless and therefore do not exist.
>
> (Hung, 2013, p. 426)

The point is not that the argument is laughably idiotic, but that it was seen at the time as a suitable rebuttal to an actual observation! We laugh now because we have the benefit of hindsight. Three centuries of the demonstrated power of the empirical approach give us an edge on poor Sizi. Take away those years of empiricism, and many of us might have been there nodding our heads and urging him on. No, the empirical approach is *not* necessarily obvious, which is why we often have to *teach* it, even in a society that is dominated by science.

Empiricism pure and simple is not enough, however. Note that the heading for this section is "*Systematic* Empiricism." Observation is fine and necessary, but pure, unstructured observation of the natural world will not lead to scientific knowledge. Write down every observation you make from the time you get up in the morning to the time you go to bed on a given day. When you finish, you will have a great number of facts, but you will not have a greater understanding of the world. Scientific observation is termed *systematic* because it is structured so that the results of the observation reveal something about the underlying nature of the world. Scientific observations are usually theory driven; they test different explanations of the nature of the world. They are structured so that, depending on the outcome of the observation, some theories are supported and others rejected.

Publicly Verifiable Knowledge: Replication and Peer Review

Scientific knowledge is public in a special sense. By *public,* we, of course, do not mean that scientific observations are posted on community center bulletin boards. Instead, we refer to the fact that scientific knowledge does not exist solely in the mind of a

particular individual. In an important sense, scientific knowledge does not exist at all until it has been submitted to the scientific community for criticism and empirical testing by others. Knowledge that is considered "special"—the province of the thought processes of a particular individual, immune from scrutiny and criticism by others—can never have the status of scientific knowledge. Likewise, science rejects the claim that particular groups have access to special knowledge (Lukianoff, 2012).

Science makes the idea of public verifiability concrete via the procedure of *replication.* In order to be considered in the realm of science, a finding must be presented to the scientific community in a way that enables other scientists to attempt the same experiment and obtain the same results. When this occurs, we say that the finding has been replicated. Scientists use replication to define the idea of public knowledge. Replication ensures that a particular finding is not due simply to the errors or biases of a particular investigator. In short, for a finding to be accepted by the scientific community, it must be possible for someone other than the original investigator to duplicate it. When a finding is presented in this way, it becomes public. It is no longer the sole possession of the original researcher; it is instead available for other investigators to extend, criticize, or apply in their own ways.

The poet John Donne told us that "no man is an island." In science, no researcher is an island. Each investigator is connected to the scientific community and its knowledge base. It is this interconnection that enables science to grow cumulatively. Researchers constantly build on previous knowledge in order to go beyond what is currently known. This process is possible only if previous knowledge is stated in such a way that any investigator can use it to build on.

By *publicly verifiable knowledge,* then, we mean findings presented to the scientific community in such a way that they can be replicated, criticized, or extended by anyone in the community. This is a most important criterion not only for scientists but also for the layperson, who, as a consumer, must evaluate scientific information presented in the media. One important way to distinguish charlatans and practitioners of pseudoscience from legitimate scientists is that the former often bypass the normal channels of scientific publication and instead go straight to the media with their "findings." One ironclad criterion that will always work for the public when presented with scientific claims of uncertain validity is the question, Have the findings been published in a recognized scientific journal that uses some type of peer review procedure? The answer to this question will almost always separate pseudoscientific claims from the real thing.

Peer review is a procedure in which each paper submitted to a research journal is critiqued by several scientists, who then submit their criticisms to an editor. The editor is usually a scientist with an extensive history of work in the specialty area covered by the journal. The editor decides whether the weight of opinion warrants publication of the paper, publication after further experimentation and statistical analysis, or rejection because the research is flawed or trivial. Legitimate journals publish statements of their editorial policies in each issue and on their websites, so you should always check to see whether a journal is peer reviewed. This is even more important now because the web has spawned dozens of open-access journals that will publish anything for a fee (Levitin, 2016). These vanity web journals prey on young scholars desperate to publish in order to get tenure at universities. Their presence on the web makes it harder for the general public to discern peer-reviewed scientific research from things on the web that look scientific but have not undergone the scrutiny of peer review.

Not all information in peer-reviewed scientific journals is necessarily correct (Gilbert et al., 2016; Open Science Collaboration, 2015), but at least it has met a criterion of peer criticism and scrutiny. Peer review is a *minimal* criterion, not a stringent one, because most scientific disciplines publish dozens of different journals of varying quality. Most scientific ideas can get published somewhere in the legitimate literature if they meet some rudimentary standards. The idea that only a narrow range of data

and theory can get published in science is false. This is an idea often suggested by purveyors of bogus remedies and therapies who try to convince the media and the public that they have been shut out of scientific outlets by a conspiracy of "orthodox science." But consider for a minute just how many legitimate outlets there are in a field like psychology. The APA database *PsycINFO* summarizes articles from over 2,000 different journals. Most of these journals are peer reviewed. Virtually all halfway legitimate theories and experiments can find their way into this vast array of publication outlets. Indeed, if anything, there are probably *too many* scientific journals. And there are *certainly* too many non-peer-reviewed journals and vanity journals that are not reviewed at all (Kolata, 2017).

Again, I am not suggesting that all ideas published in peer-reviewed psychological journals are necessarily valid. On the contrary, I emphasized earlier that this is only a minimal criterion. However, the point is that the failure of an idea, a theory, a claim, or a therapy to have adequate documentation in the peer-reviewed literature of a scientific discipline is a very sure sign. Particularly when the lack of evidence is accompanied by a media campaign to publicize the claim, *it is a sure sign that the idea, theory, or therapy is bogus.*

The mechanisms of peer review vary somewhat from discipline to discipline, but the underlying rationale is the same. Peer review is one way that science institutionalizes the attitudes of objectivity and public criticism (replication is another). Ideas and experimentation undergo a honing process in which they are submitted to other critical minds for evaluation. Ideas that survive this critical process have begun to meet the criterion of public verifiability. The peer review process is far from perfect, but it is really the only consumer protection that we have. To ignore it is to leave ourselves at the mercy of the multimillion-dollar pseudoscience industries that are so good at manipulating the media to their own ends (see Chapter 12). In subsequent chapters, we will discuss in much more detail the high price we pay for ignoring the checks and balances inherent in the true scientific practice of psychology.

Empirically Solvable Problems: Scientists' Search for Testable Theories

Science deals with solvable, or specifiable, problems. This means that the types of questions that scientists address are potentially answerable by means of currently available empirical techniques. If a problem is not solvable or a theory is not testable by the empirical techniques that scientists have at hand, then scientists will not attack it. For example, the question "Will three-year-old children given structured language stimulation during day care be ready for reading instruction at an earlier age than children not given such extra stimulation?" represents a scientific problem. It is answerable by currently available empirical methods. The question "Are human beings inherently good or inherently evil?" is not an empirical question and, thus, is simply not in the realm of science. Likewise, the question "What is the meaning of life?" is not an empirical question and so is outside the realm of science.

Science advances by positing theories to account for particular phenomena in the world, by deriving predictions from these theories, by testing the predictions empirically, and by modifying the theories based on the tests. The sequence might be portrayed as follows: theory → prediction → test → theory modification. So what a scientist often means by the term *solvable problem* is "*testable* theory." What makes a theory testable? The theory must have specific implications for observable events in the natural world; this is what is meant by *empirically testable.* This criterion of testability is often called the *falsifiability criterion*, and it is the subject of Chapter 2.

By saying that scientists tackle empirically solvable problems, we do not mean to imply that different classes of problems are inherently solvable or unsolvable and that this division is fixed forever. Quite the contrary: Some problems that are currently

unsolvable may become solvable as theory and empirical techniques become more sophisticated. For example, decades ago, historians would not have believed that the controversial issue of whether Thomas Jefferson fathered a child by his slave Sally Hemings was an empirically solvable question. Yet, by 1998, this problem had become solvable through advances in genetic technology, and a paper was published in the journal *Nature* (Foster et al., 1998) indicating that it was highly probable that Jefferson was the father of Eston Hemings Jefferson.

This is how science in general has developed and how new sciences have come into existence. There is always ample room for disagreement about what is currently solvable. Scientists themselves often disagree on this point as it relates to a particular problem. Thus, although all scientists agree on the solvability criterion, they may disagree on its specific applications. Nobel laureate Peter Medawar titled one of his books *The Art of the Soluble* to illustrate that part of the creativity involved in science is finding the problem on the farthest edge of the frontier of human knowledge that will yield to empirical techniques.

Psychology itself provides many good examples of the development from the unsolvable to the solvable. There are many questions (such as "How does a child acquire the language of his or her parents?", "Why do we forget things we once knew?", or "How does being in a group change a person's behavior and thinking?") that had been the subjects of philosophical speculation for centuries before anyone recognized that they could be addressed by empirical means. As this recognition slowly developed, psychology coalesced as a collection of problems concerning behavior in a variety of domains. Psychological issues gradually became separated from philosophy, and a separate empirical discipline evolved.

Cognitive psychologist Steven Pinker (1997) discusses how ignorance can be divided into *problems* and *mysteries.* In the case of problems, we know that an answer is possible and what that answer might look like even though we might not actually have the answer yet. In the case of mysteries, we can't even conceive of what an answer might look like. Using this terminology, we can see that science is a process that turns mysteries into problems. In fact, Pinker (1997) noted that he wrote his book *How the Mind Works* "because dozens of mysteries of the mind, from mental images to romantic love, have recently been upgraded to problems" (p. ix).

Psychology and Folk Wisdom: The Problem with "Common Sense"

We all have implicit models of behavior that govern our interactions and our thoughts about ourselves and other people. Indeed, some social, personality, and cognitive psychologists study the nature of these implicit psychological theories. However, most people never think about their theories in a clear and logical manner. Instead, we usually become aware of them only when attention is drawn to them or when we find them challenged in some way.

Actually, our personal models of behavior are not really coherent in the way that an actual theory would have to be. Instead, we carry around a ragbag of general principles, homilies, and clichés about human behavior that we draw on when we feel that we need an explanation. The problem with this commonsense knowledge, this folk wisdom, is that much of it contradicts itself and is, therefore, unfalsifiable (the principle of falsifiability is the topic of the next chapter).

Often a person uses some folk proverb to explain a behavioral event even though, on an earlier occasion, this same person used a directly contradictory folk proverb to explain the same type of event. For example, most of us have heard or said, "look before you leap." Now there's a useful, straightforward bit of behavioral

advice—except that I vaguely remember admonishing on occasion, "he who hesitates is lost." And "absence makes the heart grow fonder" is a pretty clear prediction of an emotional reaction to separation. But then what about "out of sight, out of mind"? And if "haste makes waste," why do we sometimes hear that "time waits for no man"? How could the saying "two heads are better than one" not be true? Except that "too many cooks spoil the broth." If I think "it's better to be safe than sorry," why do I *also* believe "nothing ventured, nothing gained"? And if "opposites attract," why do "birds of a feather flock together"? I have counseled many students to "never to put off until tomorrow what you can do today." But I hope my last advisee has never heard me say this, because I just told him, "cross that bridge when you come to it."

The enormous appeal of clichés like these is that, taken together as implicit "explanations" of behavior, they cannot be refuted. No matter what happens, one of these explanations will be cited to cover it. No wonder we all think we are such excellent judges of human behavior and personality. Our folk wisdom gives us an explanation for anything and everything that happens. As such, folk wisdom is cowardly in the sense that it takes no risk that it might be refuted.

That folk wisdom is "after the fact" wisdom, and that it actually is useless in a truly predictive sense, is why sociologist Duncan Watts titled one of his books: *Everything Is Obvious—Once You Know the Answer* (2011). Watts discusses a classic paper by Lazarsfeld (1949) in which, over 60 years ago, he was dealing with the common criticism that "social science doesn't tell us anything that we don't already know." Lazarsfeld listed a series of findings from a massive survey of 600,000 soldiers who had served during World War II; for example, that men from rural backgrounds were in better spirits during their time of service than soldiers from city backgrounds. People tend to find all of the survey results to be pretty obvious. In this example, for instance, people tend to think it obvious that rural men would have been used to harsher physical conditions and thus would have adapted better to the conditions of military life. It is likewise with all of the other findings—people find them pretty obvious. Lazarsfeld then reveals his punchline: All of the findings were the opposite of what was originally stated. For example, it was *actually* the case that men from city backgrounds were in better spirits during their time of service than soldiers from rural backgrounds. The last part of the learning exercise is for people to realize how easily they would have explained just the opposite finding. In the case of the actual outcome, people tend to explain it (when told of it first) by saying that they expected it because city men are used to working in crowded conditions and under hierarchical authority. They never realize how easily they would have concocted an explanation for exactly the opposite finding!

So sometimes our folk theories of behavior can't be refuted. We will see in the next chapter why this inability to be refuted makes such theories not very useful. However, a further problem occurs even in cases in which our folk beliefs do have some specificity; that is, even when they are empirically testable. The problem is that psychological research has shown that, when many common cultural beliefs about behavior are subjected to empirical test, they turn out to be false.

It is not difficult to generate instances of folk beliefs (or "common sense") that are wrong. Take, for example, the idea that children who excel academically, or who read a lot, are not socially or physically adept. This idea still circulates in our society even though it is utterly false. There is evidence that, contrary to "commonsense" folk belief, readers and academically inclined individuals are more physically robust and are more socially involved than are people who do not read (Singh et al., 2012). People who are readers are more likely to play sports, jog, exercise, and participate in extracurricular activities than are people who do not read very much.

Many of our folk beliefs about behavior arise and take on a life of their own. For example, over several years, the folk belief has developed in our society and in schools that low self-esteem is a cause of aggression. But empirical investigations

schools that low self-esteem is a cause of aggression. But empirical investigations

have indicated that there is no connection between aggression and low self-esteem (Bushman et al., 2009; Krueger et al., 2008). If anything, the opposite appears to be the case—aggression is more often associated with high self-esteem. Likewise, an extremely popular hypothesis for the past couple of decades has been that school achievement problems are the result of low self-esteem in students. In fact, it turns out that the relationship between self-esteem and school achievement is more likely to be in the opposite direction from that assumed by educators and parents. It is superior accomplishment in school (and in other aspects of life) that leads to high self-esteem and not the reverse (Lilienfeld et al., 2012).

Another example of folk wisdom gone wrong is the common admonition to students that if they become unsure about an answer that they have given on a multiple-choice test they should never switch from their original choice. Not only do most students think that they should not switch when uncertain of an answer, but even Barron's Guide to GRE Preparation advises "Exercise great caution if you decide to change an answer. Experience indicates that many students who change answers change to the wrong answer" (Kruger et al., 2005, p. 725). This advice is completely wrong. The advice is wrong because the folk myth that changing answers decreases a person's score is dead wrong. Actual research has shown that when doubts about a multiple-choice answer arise, students are better off switching from their first answer (Kruger et al., 2005; Lilienfeld et al., 2010).

A case where we can really see folk wisdom run amok is in the folk myth that we use only 10 percent of our brainpower. Despite having absolutely no basis in cognitive neuroscience (Ferguson et al., 2017; Lilienfeld et al., 2010; Stix, 2015), this one has been around for decades and has taken on the status of what has been termed a "psycho-fact"—a statement about psychology that is not true but which has been repeated so much that the average person thinks that it is a fact. It is likewise with the belief that some people are "left-brained" and other people are "right-brained," or that certain aspects of personality are controlled by the left side of the brain and other aspects of personality by the right side. Although modern neuroscience research does show subtle specializations throughout the brain, the strongly stated popularizations of this idea in terms of "left" or "right" are invariably nonsense—particularly in the context of the finding that our brains work in an integrated fashion (Lilienfeld et al., 2010; Radford, 2011).

Consider what happened at the trial of former White House aide Lewis (Scooter) Libby in 2007. Expert testimony from a renowned research psychologist was disallowed because the judge ruled that it was well known that memory was fallible and that juries can safely rely on their common sense to ascertain how memory works. In fact, studies show that almost 30 percent of the population believe that human memory "operates like a tape recorder" (Lilienfeld, 2012). Contrary to what the judge thought, 30 percent of his jury badly needed to hear from the expert!

Folk beliefs are not always immune to evidence. Sometimes, when the contradictory evidence becomes too widely known, folk psychology ("common sense") does change. For example, years ago, one widely held cliché about children was "Early ripe, early rot." The cliché reflected the belief that childhood precocity was associated with adult abnormality, a belief sustained by many anecdotes about childhood prodigies who came to ruin in later life. In this case, the psychological evidence documenting the inaccuracy of the cliché has been absorbed into the general culture, and you will almost never hear this bit of folk "wisdom" anymore.

Nevertheless, new folk myths are being created all the time, and a few will, of course, go viral. It seems that for every folk myth that we kill with evidence, another pops up in its place! For example, in the past decade it has been common to hear the claim that the millennial generation, because it has grown up with technology, has the ability to multitask—that they can add secondary tasks to the primary one they are

doing without any loss of efficiency in accomplishing the primary task. Millennials

themselves sometimes claim that they can study just as well while texting friends and watching television at the same time. This folk belief is false. Millennials are no better than anyone else at multitasking, because research shows that virtually all people are performance-impaired when multitasking (Jaffe, 2012; Kirschner & van Merrienboer, 2013; Ophir et al., 2009; Strayer et al., 2013). When doing additional tasks, *everyone's* performance (millennials and nonmillennials) on their primary task is impeded.

The same wishful thinking that has fueled the idea that multitasking is possible facilitates the development of other folk myths. For example, many people believe that "we are each intelligent in our own way"—a belief not supported by actual research on the nature of intelligence (Deary, 2013; Ferguson et al., 2017; Waterhouse, 2006). Other people believe that speed reading (reading at several times the normal rate with little loss in comprehension) is possible. It is not (Rayner et al., 2016).

These problems with folk psychology would not be so damaging if people realized the fallibility of their folk beliefs. Instead, however, surveys have shown (see Lilienfeld, 2012) that over 80 percent of the public thinks that adequate training in psychology is provided by daily life! To the contrary, we need the discipline of psychology because it provides tests of the empirical basis of common sense. Sometimes common sense beliefs do not hold up when tested, as we saw in many of the previous examples. From the examples discussed—and many more could be cited—we can see that psychology's role as the empirical tester of much folk wisdom often brings it into conflict with many widely held cultural beliefs. Psychology is often the bearer of the "bad tidings" that comfortable folk beliefs do not stand up to the cold light of day. Perhaps it is not surprising that many people would like not only to ignore the message but also to do away with the messenger.

Psychology as a Young Science

Psychology's battle to establish its problems as empirically solvable has only recently been won. But as the science progresses, psychologists will address more and more issues that are the subject of strongly held beliefs about human beings because many of these problems are empirically testable. Psychologists now study such highly charged topics as the development of moral reasoning, the psychology of romantic love, the efficacy of prayer, the determinants of crime, the efficacy of different family structures, and the factors that make divorce more likely. Studies of childhood sexual activity, for example, have incited much controversy (Lilienfeld, 2010; Rind, 2008). Psychology studies many things that people have strong opinions about, such as altruism, greed, and lying (Ariely, 2013; Greene, 2013). Some people object to empirical investigation in these areas; yet there has been scientific progress in each one of them.

Finally, people often are offended simply by the presentation of simple descriptive facts about human behavior. For example, it is enough to offend some people to just report the simple fact that children growing up in single-parent households are more likely to experience poverty and behavioral problems (Chetty et al., 2014; McLanahan et al., 2013; Murray, 2012; Parker, 2014). This is the type of opposition to simple empirical facts about human behavior that psychology has to deal with on a regular basis. This opposition, when the issue is a heated one, can get hostile and can be personally directed at psychologists. Memory researcher Elizabeth Loftus was subjected to death threats and lawsuits when her research revealed that people's claims of having uncovered repressed memories of abuse and molestation were not reflective of reality (Observations, 2017; Sleek, 2017). There are subgroups in the population who do not like many of the findings coming out of scientific psychology, including, for example, that intelligence is partially heritable; that there are evolutionary explanations for some of our sexual behaviors; and that there are cognitive biases that lead us to believe

in pseudoscience and conspiracy theories (Buss, 2011; Deary, 2013; Ferguson, et al., 2017; Shermer 2011, 2017; Stanovich et al., 2016).

Psychology is often in a no-win situation as a discipline. On the one hand, some people object to calling psychology a science and deny that psychologists can establish empirical facts about behavior. On the other hand, there are those who object to the investigation of certain areas of human behavior because they fear that the facts uncovered by psychology might threaten their beliefs. Skinnerian psychologists regularly deal with these contradictory criticisms. For instance, critics have argued that the laws of reinforcement formulated by behaviorists do not apply to human behavior. At the same time, other critics are concerned that the laws will be used for the rigid and inhumane control of people. Thus, the behaviorists are faced with some critics who deny that their laws can be applied and others who charge that their laws can be applied all too easily!

Examples such as this arise because the relatively new science of psychology has just begun to uncover facts about aspects of behavior that have previously escaped study. The relative youth of psychology as a science partially explains why many people are confused about the discipline. Nevertheless, during the past several decades, psychology has become firmly established in the interconnecting structure of knowledge that we call science. Failure to appreciate this fact is the source of almost all of the confused thinking about psychology that you will encounter.

Summary

Psychology is an immensely diverse discipline covering a range of subjects that are not always tied together by common concepts. Instead, what unifies the discipline is that it uses scientific methods to understand behavior. The scientific method is not a strict set of rules; instead it is defined by some very general principles. Three of the most important are that (1) science employs methods of systematic empiricism; (2) it aims for knowledge that is publicly verifiable; and (3) it seeks problems that are empirically solvable and that yield testable theories (the subject of the next chapter). The structured and controlled observations that define systematic empiricism are the subject of several later chapters of this book. Science renders knowledge public by procedures such as peer review and mechanisms such as replication.

Psychology is a young science and, thus, is often in conflict with so-called folk wisdom. This conflict is typical of all new sciences, but understanding it helps to explain some of the hostility directed toward psychology as a discipline. This characteristic of questioning common wisdom also makes psychology an exciting field. Many people are drawn to the discipline because it holds out the possibility of actually testing “common sense” that has been accepted without question for centuries.

Chapter 2

Falsifiability: How to Foil Little Green Men in the Head

Learning Objectives

2.1 Illustrate the importance of falsifiability to scientific theory

2.2 Describe how, when a science advances, the errors made increase in specificity

In 1793, a severe epidemic of yellow fever struck Philadelphia. One of the leading doctors in the city at the time was Benjamin Rush, a signer of the Declaration of Independence. During the outbreak, Rush was one of the few physicians who were available to treat literally thousands of yellow fever cases. Rush adhered to a theory of medicine that dictated that illnesses accompanied by fever should be treated by vigorous bloodletting (the removal of blood from the body either by using an instrument such as a lancet or by the application of leeches). He administered this treatment to many patients, including himself when he came down with the illness. Critics charged that his treatments were more dangerous than the disease. However, following the epidemic, Rush became even more confident of the effectiveness of his treatment, even though many of his patients had died. Why?

One writer summarized Rush's attitude this way: "Convinced of the correctness of his theory of medicine and lacking a means for the systematic study of treatment outcome, he attributed each new instance of improvement to the efficacy of his treatment and each new death that occurred despite it to the severity of the disease" (Eisenberg, 1977, p. 1106). In other words, if the patient got better, this improvement was taken as proof that bloodletting worked. If instead the patient died, Rush interpreted this to mean that the patient had been too ill for *any* treatment to work. We now know that Rush's critics were right: His treatments were as dangerous as the disease. In this chapter, we will discuss how Rush went wrong. His error illustrates one of the most important principles of scientific thinking, one that is particularly useful in evaluating psychological claims.

In this chapter, we focus in more detail on the third general characteristic of science that we discussed in Chapter 1: Scientists deal with solvable problems. What scientists most often mean by a *solvable problem* is a "testable theory." The way scientists make sure they are dealing with testable theories is by ensuring that their theories are falsifiable, that is, that they have implications for actual events in the natural world. We will see why what is called *the falsifiability criterion* is so important in psychology.

Theories and the Falsifiability Criterion

Benjamin Rush fell into a fatal trap when assessing the outcome of his treatment. His method of evaluating the evidence made it impossible to conclude that his treatment did not work. If the recovery of a patient meant confirmation of his treatment (and, hence, his theory of medicine), then it only seems fair that the death of a patient should have meant disconfirmation. Instead, he rationalized away these disconfirmations. By interpreting the evidence as he did, Rush violated one of the most important rules regarding the construction and testing of theories in science: He made it impossible to falsify his theory.

Scientific theories must always be stated in such a way that the predictions derived from them could potentially be shown to be false (Koepsell, 2015). Thus, the methods of evaluating new evidence relevant to a particular theory must always include the possibility that the data will falsify the theory. This principle is often termed the *falsifiability criterion*, and its importance in scientific progress has been most forcefully articulated by Karl Popper, a philosopher of science whose writings are read widely by working scientists (Firestein, 2016).

The falsifiability criterion states that, for a theory to be useful, the predictions drawn from it must be specific. The theory must go out on a limb, so to speak, because in telling us what *should* happen, the theory must also imply that certain things will *not* happen. If these latter things *do* happen, then we have a clear signal that something is wrong with the theory: It may need to be modified, or we may need to look for an entirely new theory. Either way, we will end up with a theory that is nearer to the truth. By contrast, if a theory does not rule out any possible observations, then the theory can never be changed, and we are frozen into our current way of thinking, with no possibility of progress. Thus, a successful theory is not one that accounts for every possible outcome because such a theory robs itself of any predictive power. As biologist Stuart Firestein (2016) puts it, we should have confidence in science not because it is always right, but instead because it is possible to prove it wrong.

Because we shall often refer to the evaluation of theories in the remainder of this book, we must clear up one common misconception surrounding the word *theory.* The misconception is reflected in the commonly used phrase "Oh, it's only a theory." This phrase captures what *laypeople* often mean when they use the word *theory:* an unverified hypothesis, a mere guess, a hunch. It implies that one theory is as good as another. This is most definitely *not* the way the word *theory* is used in science! When scientists refer to theories, they do not mean unverified guesses.

A theory in science is an interrelated set of concepts that is used to explain a body of data and to make predictions about the results of future experiments. *Hypotheses* are specific predictions that are derived from theories (which are more general and comprehensive). Currently viable theories are those that have had many of their hypotheses confirmed. The theoretical structures of such theories are, thus, consistent with a large number of observations. However, when the database begins to contradict the hypotheses derived from a theory, scientists begin trying to construct a new theory (or, more often, simply make adjustments in the previous theory) that will provide a better interpretation of the data. Thus, the theories that are under scientific discussion are those that have been verified to some extent and that do not make many predictions that are contradicted by the available data. They are *not* mere guesses or hunches.

The difference between the layperson's and the scientist's use of the word *theory* has often been exploited by people who want creationism taught in the public schools (Miller, 2008; Scott, 2005). Their argument often is "After all, evolution is only a theory." This statement is intended to suggest the *layperson's* use of the term *theory.* In common language, the term theory means "only a guess." However, the theory of evolution by natural selection is not a theory in the layperson's sense (to the contrary,

in the layperson's sense, it would be called a fact; see Randall, 2005). Instead, it is a

theory in the *scientific* sense. It is a conceptual structure that is supported by a large and varied set of data (Dawkins, 2010, 2016). It is not a mere guess, equal to any other guess. Instead, it interlocks with knowledge in a host of other disciplines, including geology, physics, chemistry, and all aspects of biology. The distinguished biologist Theodosius Dobzhansky (1973) made this point in a famous article titled "Nothing in Biology Makes Sense Except in the Light of Evolution."

The Theory of Knocking Rhythms

A hypothetical example will show how the falsifiability criterion works. A student knocks at my door. A colleague in my office with me has a theory that makes predictions about the rhythms that different types of people use to knock. Before I open the door, my colleague predicts that the person behind it is a female. I open the door and, indeed, the student is a female. Later I tell my colleague that I am impressed, but only mildly so because he had a 50 percent chance of being correct even without his "theory of knocking rhythms"—actually even higher, because on most campuses, females outnumber males. He says he can do better. Another knock comes. My colleague tells me it is a male under 22 years old. I open the door to find a male student whom I know to be just out of high school. I comment that I am somewhat impressed because our university has a considerable number of students over the age of 22. Yet I still maintain that, of course, young males are quite common on campus. Thinking me hard to please, my colleague proposes one last test. After the next knock, my colleague predicts, "Female, 30 years old, 5 feet 2 inches tall, carrying a book and a purse in the left hand and knocking with the right." After opening the door and confirming the prediction completely, I have quite a different response. I say that, assuming my colleague did not play a trick and arrange for these people to appear at my door, I am now in fact extremely impressed.

Why the difference in my reactions? Why do my friend's three predictions yield three different responses, ranging from "So what?" to "Wow!"? The answer has to do with the specificity and precision of the predictions. The more specific predictions made a greater impact when they were confirmed. Notice, however, that the specificity varied directly with the falsifiability. The more specific and precise the prediction was, the more potential observations there were that could have falsified it. For example, there are a lot of people who are *not* 30-year-old females and 5 feet 2 inches tall. Note that implicitly, by my varied reactions, I signaled that I would be more impressed by a theory that made predictions that maximized the number of events that should *not* occur.

Good theories, then, make predictions that expose themselves to falsification. Bad theories do not put themselves in jeopardy in this way. They make predictions that are so general that they are almost bound to be true (e.g., the next person to knock on my door will be less than 100 years old) or are phrased in such a way that they are completely protected from falsification (as in the Benjamin Rush example). In fact, a theory can be so protected from falsifiability that it is simply no longer considered scientific at all. Indeed, it was philosopher Karl Popper's attempt to define the criteria that separate science from nonscience that led him to emphasize the importance of the falsifiability principle. There is a direct link here to psychology and to our discussion of Freud in Chapter 1.

Freud and Falsifiability

In the early decades of the twentieth century, Popper was searching for the underlying reasons that some scientific theories seem to lead to advances in knowledge and others lead to intellectual stagnation (Hacohen, 2000). Einstein's general relativity theory, for example, led to startlingly new observations (for instance, that the light from a

distant star bends when it passes near the sun) precisely because its predictions were

structured so that many possible events could have contradicted them and, thus, falsified the theory (Firestein, 2016).

Popper reasoned that this is not true of stagnant theories—and pointed to Freudian psychoanalysis as an example. Freudian theory uses a complicated conceptual structure that explains human behavior after the fact—that is, after it has occurred—but does not predict things in advance. In short, Freudian theory can explain everything. However, as Popper argued, it is precisely this property that makes it scientifically useless. It makes no specific predictions. Adherents of psychoanalytic theory spend much time and effort in getting the theory to explain every known human event, from individual quirks of behavior to large-scale social phenomena. But their success in making the theory a rich source of after-the-fact explanation robs it of any scientific utility. Freudian psychoanalytic theory currently plays a much larger role as a spur to the literary imagination than as a theory in contemporary psychology. Its demise within psychology can be traced in part to its failure to satisfy the falsifiability criterion (Boudry & Buekens, 2011).

But the existence of such unfalsifiable theories does real damage. For example, explanations for the cause of autism (in part a genetically determined disorder) were led down a blind alley by psychoanalytic explanations for the condition. Influenced by psychoanalytic ideas, psychologist Bruno Bettelheim popularized the now-discredited notion of "refrigerator mothers" as the cause and thought that "the precipitating factor in infantile autism is the parent's wish that his child should not exist" (Offit, P. A. 2008, p. 3). Ideas like this not only did damage, but they set back the study of autism.

As another example, consider the history of Gilles de la Tourette syndrome. This is a disorder characterized by physical tics and twitches that may involve any part of the body, as well as vocal symptoms such as grunts and barks, echolalia (involuntary repetition of the words of others), and coprolalia (compulsive repetition of obscene words). Tourette syndrome is an organically based disorder of the central nervous system and is now often successfully treated with drug therapies (Smith et al., 2016; Suski & Stacy, 2016). Importantly, understanding of the cause and treatment of the disorder was considerably hampered from 1921 to 1955, when explanations and treatments for Tourette syndrome were dominated by psychoanalytic conceptualizations (see Kushner, 1999). Author after author presented unfalsifiable psychoanalytic explanations for the syndrome. The resulting array of vague explanations created a conceptual sludge that obscured the true nature of the syndrome and probably impeded scientific progress toward an accurate understanding of it. For example, according to one author,

> [Tourette syndrome is] a classic example of the retrogressive effect of psychoanalysis on the investigation of brain disease. La Tourette had attributed the disease to a degenerative process of the brain. After Freud's theories became fashionable in the early decades of the twentieth century, attention in such conditions was deflected from the brain. . . . The consequence of this retrograde movement was that patients tended to be referred to psychiatrists (usually of a psychoanalytic persuasion) rather than to neurologists, so that physical examinations and investigation were not performed.
>
> (Thornton, 1986, p. 210)

Shapiro et al. (1978) described one psychoanalyst who thought that his patient was "reluctant to give up the tic because it became a source of erotic pleasure to her." A second considered the tic a "conversion symptom at the anal-sadistic level." A third thought that a person with Tourette syndrome had a "compulsive character, as well as a narcissistic orientation" and that the patient's tics "represent[ed] an affective syndrome, a defense against the intended affect." Psychologist Jerome Kagan (2006) tells us how Sandor Ferenczi, a disciple of Freud *who had never seen a patient with Tourette's syndrome* wrote that "the frequent facial tics of people with Tourette's were the result of

a repressed urge to masturbate" (p. 179).

Progress in the treatment and understanding of Tourette syndrome began to occur only when researchers recognized that the psychoanalytic "explanations" were useless. These explanations were enticing because they seemed to explain things. In fact, they explained everything—after the fact. However, the explanations they provided created only the illusion of understanding. By attempting to explain everything after the fact, they barred the door to any advancement. Progress occurs only when a theory does not predict *everything* but instead makes specific predictions that tell us—in advance—something specific about the world. The predictions derived from such a theory may be wrong, of course, but this is a strength, not a weakness.

The Little Green Men

It is not difficult to recognize unfalsifiable conceptualizations when one is detached from the subject matter and particularly when one has the benefit of historical hindsight (as in the Benjamin Rush example). It is also easy to detect unfalsifiable conceptualizations when the instance is obviously concocted. For example, it is a little known fact that I have discovered the underlying brain mechanism that controls behavior. You will soon be reading about this discovery (in the *National Enquirer*, available at your local supermarket). In the left hemisphere of the brain, near the language areas, reside two tiny green men. They have the power to control the electrochemical processes taking place in many areas of the brain. And, well, to make a long story short, they basically control everything. There is one difficulty, however. The little green men have the ability to detect any intrusion into the brain (surgery, X-rays, etc.), and when they do sense such an intrusion, they tend to disappear. (I forgot to mention that they have the power to become invisible.)

I have no doubt insulted your intelligence by using an example more suitable for elementary school students. I have obviously concocted this example so that my hypothesis about the little green men could never be shown to be wrong. However, consider this. As a lecturer and public speaker on psychological topics, I am often confronted by people who ask me why I have not lectured on all the startling new discoveries in extrasensory perception (ESP) and parapsychology that have been made in the past few years. I have to inform these questioners that most of what they have heard about these subjects has undoubtedly come from the general media rather than from scientifically respectable sources. In fact, some scientists have looked at these claims and have not been able to replicate the findings. I remind the audience that replication of a finding is critical to its acceptance as an established scientific fact and that this is particularly true in the case of results that contradict either previous data or established theory.

I further admit that many scientists have lost patience with ESP research. Although one reason is undoubtedly that the area is tainted by fraud, charlatanism, and media exploitation, perhaps the most important reason for scientific disenchantment is the existence of what science writer Martin Gardner years ago called the catch-22 of ESP research.

It works as follows: A "believer" (someone who believes in the existence of ESP phenomena before beginning an investigation) claims to have demonstrated ESP in the laboratory. A "skeptic" (someone who doubts the existence of ESP) is brought in to confirm the phenomena. Often, after observing the experimental situation, the skeptic calls for more controls (controls of the type we will discuss in Chapter 6), and though these are sometimes resisted, well-intentioned believers often agree to them. When the controls are instituted, the phenomena cannot be demonstrated (Galak et al., 2012; Hand, 2014; Nickell & McGaha, 2015). The skeptic, who correctly interprets this failure as an indication that the original demonstration was due to inadequate experimen-

tal control and, thus, cannot be accepted, is often shocked to find that the believer does not regard the original demonstration as invalid. Instead, the believer invokes

the catch-22 of ESP: Psychic powers, the believer maintains, are subtle, delicate, and easily disturbed. The "negative vibes" of the skeptic were probably responsible for the disruption of the "psi powers." The believer thinks that the powers will undoubtedly return when the negative aura of the skeptic is removed.

This way of interpreting failures to demonstrate ESP in experiments is logically analogous to my story about the little green men. ESP operates just as the little green men do. It's there as long as you don't intrude to look at it carefully. When you do, it disappears. If we accept this explanation, it will be impossible to demonstrate the phenomenon to any skeptical observers. It appears only to believers. Of course, this position is unacceptable in science. We do not have the magnetism physicists and the nonmagnetism physicists (those for whom magnetism does and does not "work"). Interpreting ESP experiments in this way makes the hypothesis of ESP unfalsifiable just as the hypothesis of the little green men is. Interpreting the outcomes in this way puts it outside the realm of science.

Not All Confirmations Are Equal

The principle of falsifiability has important implications for the way we view the confirmation of a theory. It is too simple to think that it is only the *amount* of confirming evidence that is critical in the evaluation of a theory. Falsifiability implies that the number of *times* a theory has been confirmed is not the critical element. The reason is that, as our example of the "theory of knocking rhythms" illustrated, not all confirmations are equal. Confirmations are more or less impressive depending on the extent to which the prediction exposes itself to potential disconfirmation. One confirmation of a highly specific, potentially falsifiable prediction (for instance, a female, 30 years old, 5 feet 2 inches tall, carrying a book and a purse in the left hand and knocking with the right) has a greater impact than the confirmation of 20 different predictions that are all virtually unfalsifiable (for instance, a person less than 100 years old).

Thus, we must look not only at the *quantity* of the confirming evidence, but also at the *quality* of the confirming instances. Using the falsifiability criterion as a tool to evaluate evidence will help the research consumer resist the allure of the nonscientific, all-explaining theory that inevitably hinders the search for deeper understanding. Indeed, such theoretical dead ends are often tempting precisely because they can never be falsified. They are islands of stability in the chaotic modern world.

Popper often made the point that "the secret of the enormous psychological appeal of these unfalsifiable theories lies in their ability to explain everything. To know in advance that whatever happens you will be able to understand it gives you not only a sense of intellectual mastery but, even more important, an emotional sense of secure orientation in the world" (Magee, 1985, p. 43). However, the attainment of such security is not the goal of science, because such security would be purchased at the cost of intellectual stagnation. Science is a mechanism for continually challenging previously held beliefs by subjecting them to empirical tests in such a way that they can be shown to be wrong. This characteristic often puts science—particularly psychology—in conflict with so-called folk wisdom or common sense (as we discussed in Chapter 1).

Falsifiability and Folk Wisdom

Psychology is a threat to the comfort that folk wisdom provides because, as a science, it cannot be content with explanations that cannot be refuted. The goal of psychology is the empirical testing of alternative behavioral theories in order to rule out some of them. Aspects of folk wisdom that are explicitly stated and that do stand up to empirical testing are, of course, welcomed, and many have been incorporated into psychological theory. However, psychology does not seek the comfort of explanatory systems

that account for everything after the fact but predict nothing in advance. It does not accept systems of folk wisdom that are designed never to be changed and that end

up being passed on from generation to generation. It is self-defeating to try to hide this fact from students or the public. Unfortunately, some psychology instructors and popularizers are aware that psychology's threat to folk wisdom disturbs some people, and they sometimes seek to soothe such feelings by sending a false underlying message that implies, "You'll learn some interesting things, but don't worry—psychology won't challenge things you believe in strongly." This is a mistake, and it contributes to confusion both about what science is and about what psychology is. Psychology establishes facts about sexual behavior, intelligence, crime, financial behavior, the effects of marriage, child rearing, and many other topics that people feel strongly about. It would be amazing if the investigation of subjects such as these failed to uncover something that did not upset *somebody*!

Science *seeks* conceptual change. Scientists try to describe the world as it really is, as opposed to what our prior beliefs dictate it should be. The dangerous trend in modern thought is the idea that people must be shielded from the nature of the world—that a veil of ignorance is necessary to protect a public unequipped to deal with the truth. Psychology is like other sciences in rejecting the idea that people need to be shielded from findings that make them uncomfortable. Psychology is not a "safe space" for those who want their beliefs to go unchallenged by evidence (Furedi, 2017; Lilienfeld, 2017; Lukianoff & Haidt, 2015).

The Freedom to Admit a Mistake

Scientists have found that one of the most liberating and useful implications of the falsifiability principle is that, in science, making a mistake is not a sin. Philosopher Daniel Dennett (1995) has said that the essence of science is "making mistakes in public—making mistakes for all to see, in the hopes of getting the others to help with the corrections" (p. 380). By the process of continually adjusting theory when data do not accord with it, scientists collectively arrive at theories that better reflect the nature of the world. Biologist Stuart Firestein (2016) writes that the usual list of the pillars of science—things like reason and fact and truth and experiment and objectivity—usually have one critical pillar missing. That pillar that we often forget, Firestein suggests, is failure. By failure, Firestein means making errors that we learn something from. He mean errors in the Popperian sense. Indeed, Firestein calls Popper the philosopher of failure.

In fact, our way of operating in everyday life might be greatly improved if we could use the falsifiability principle on a personal level. This is why the word *liberating* was used in the opening sentence of this section. It has a personal connotation that was specifically intended—because the ideas developed here have implications beyond science. We would have many fewer social and personal problems if we could only understand that, when our beliefs are contradicted by evidence in the world, it is better to adjust our beliefs than to deny the evidence and cling tenaciously to dysfunctional ideas.

How many times have you been in an intense argument with someone when right in the middle—perhaps just as you were giving a heated reply and defending your point of view—you realized that you were wrong about some critical fact or piece of evidence? What did you do? Did you back down and admit to the other person that you had assumed something that wasn't true and that the other person's interpretation now seemed more correct to you? Probably not. If you are like most of us, you engaged in endless rationalization. You tried to extricate yourself from the argument without admitting defeat. The last thing you would have done was admit that you were wrong. Thus, both you and your partner in the argument became a little more confused about which beliefs more closely tracked the truth. If refutations never became public (as they do in

science), if both true and false beliefs were defended with equal vehemence, and if the correct feedback about the effects of argument was not given (as in this example), there

is no mechanism for getting beliefs more reliably in sync with reality. This is why so much of our private and public discourse is confused and why the science of psychology is a more reliable guide to the causes of behavior than is so-called common sense.

Many scientists have attested to the importance of understanding that making errors in the course of science is normal and that the real danger to scientific progress is our natural human tendency to avoid exposing our beliefs to situations in which they might be shown to be wrong. Scientists must avoid this tendency, and Nobel Prize winner Peter Medawar (1979) urged them to avoid it by remembering that "*the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not*" (p. 39; italics in original).

Here is a way of thinking about what Medawar is saying. On his show on October 17, 2005, comedian Stephen Colbert coined the term "truthiness" (Zimmer, 2010). Truthiness is the "quality of a thing *feeling* true without any evidence suggesting it actually was" (Manjoo, 2008, p. 189). What Medawar is saying is that science rejects truthiness. This often puts science at odds with modern society, where truthiness is more prevalent than ever.

Many of the most renowned scientists in psychology have followed Medawar's advice: "the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not." In an article on the career of noted experimental psychologist Robert Crowder, one of his colleagues, Mahzarin Banaji, is quoted as saying that "he is the least defensive scientist I know. If you found a way to show that his theory was wobbly or that his experimental finding was limited or flawed, Bob would beam with pleasure and plan the demise of his theory with you" (Azar, 1999, p. 18). Azar (1999) describes how Crowder developed a theory of one component of memory called precategorical acoustic storage and then carefully designed the studies that falsified his own model.

But the falsifying attitude doesn't always have to characterize each and every scientist for *science* to work. The unique power of science to reveal knowledge about the world does *not* arise because scientists are uniquely virtuous (that they are completely objective; that they are never biased in interpreting findings, etc.) but instead it arises because fallible scientists are immersed in a process of checks and balances—in a process in which other scientists are always there to criticize and to root out the errors of their peers. The strength of science comes not because scientists are especially virtuous, but from a social process where scientists constantly cross-check each others' knowledge and conclusions.

This social cross-checking is a really distinctive feature of science as a domain. Yes, the value of objectivity is often invoked by many people in other domains of life. But no other domain of life has such structured cross-checking built in like science. Instead, in other domains of life, myside bias reigns supreme. Myside bias is the tendency for people to evaluate evidence, generate evidence, and test hypotheses in a manner biased toward their own prior beliefs, opinions and attitudes (Stanovich, West, & Toplak, 2013). The farcical newspaper, *The Onion*, caricatured myside bias (Paulos, 2016) with the headline: "Majority of Parents Abuse Children, Children Report!"

Myside bias is common in the domain of politics. For example, liberals routinely admonish conservatives for failing to acknowledge the strong evidence that human activity is a contributor to global warming. They are right to do so because, as will be discussed in Chapter 8, the evidence for this conclusion is highly convergent. But what most liberals fail to realize is that they have fallen prey to myside bias by cherry-picking the issue of climate change. They have focused on an issue where it is easy for liberals to agree with the scientific conclusion and hard for conservatives to agree with the conclusion. What liberals goading conservatives about climate change do not seem to realize is how easily the situation could be cherry-picked in the other direction. They don't realize how easy it would be for conservatives to ask liberals to accept the scientific evidence for conclusions that make liberals uncomfortable—that intelligence is at least moderately heritable for example, or that women do not

in fact make 20 percent less than a man makes for doing the same job (see Chapters 6 and 12). Both liberals and conservatives are guilty of myside bias. That is why we

need science—for its *process*, not for its individuals. The latter are no more unbiased than anyone else, but they are immersed in a process of error detection and cross-checking that is relatively unique.

Thoughts Are Cheap

Our earlier discussion of the idea of testing folk wisdom leads us to another interesting corollary of the falsifiability principle: Thoughts are cheap. To be specific, what we mean here is that certain *kinds* of thoughts are cheap. Biologist and science writer Stephen J. Gould (1987) illustrated this point:

> Fifteen years of monthly columns have brought me an enormous correspondence from nonprofessionals about all aspects of science. . . . I have found that one common misconception surpasses all others. People will write, telling me that they have developed a revolutionary theory, one that will expand the boundaries of science. These theories, usually described in several pages of single-spaced typescript, are speculations about the deepest ultimate questions we can ask—what is the nature of life? the origin of the universe? the beginning of time? But thoughts are cheap. Any person of intelligence can devise his half dozen before breakfast. Scientists can also spin out ideas about ultimates. We don't (or, rather, we confine them to our private thoughts) because we cannot devise ways to test them, to decide whether they are right or wrong. What good to science is a lovely idea that cannot, as a matter of principle, ever be affirmed or denied? (p. 18)

The answer to Gould's last question is "No good at all." The types of thoughts that Gould is saying are cheap and are those that we referred to in our discussion of Karl Popper's views: grand theories that are so global, complicated, and "fuzzy" that they can be used to explain anything—theories constructed more for emotional support because they are not meant to be changed or discarded. Gould was telling us that such theories are useless for scientific purposes, however comforting they may be. Science is a creative endeavor, but the creativity involves getting conceptual structures to fit the confines of empirical data. This is tough. These types of thoughts—those that explain the world as it *actually is*—are not cheap. Probably this is why good scientific theories are so hard to come by and why unfalsifiable pseudoscientific belief systems proliferate everywhere—the latter are vastly easier to construct.

In fact, they are so easy to construct that there is an event—the Festival of Bad Ad-Hoc Hypotheses (BAH-Fest)—that gives an award for presenting the most creative theory that is made unfalsifiable by ad hoc assumptions and caveats (Chen, 2014). One reason that constructing unfalsifiable theories is easy is that one sure-fire way to prevent a theory from being falsified is to fill it with obfuscation and incomprehensible jargon (Davies, 2012). Many unfalsifiable conspiracy theories (e.g., that the government has spread AIDS deliberately, or that it knew of the 9/11 attacks in advance) have this property. This is the reason that belief in conspiracy theories tends to correlate—if a person believes in one, they tend to believe in another (Lewandowsky et al., 2013; Majima, 2015). There seems to be a general tendency to be entranced by unfalsifiable obfuscation.

Errors in Science: Getting Closer to the Truth

In the context of explaining the principle of falsifiability, we have outlined a simple model of scientific progress. Theories are put forth and hypotheses are derived from

them. The hypotheses are tested by a variety of techniques that we shall discuss in the remainder of this book. If the hypotheses are confirmed by the experiments,

then the theory receives some degree of corroboration. If the hypotheses are falsified by the experiments, then the theory must be altered in some way, or it must be discarded for a better theory.

Of course, saying that knowledge in science is tentative and that hypotheses derived from theories are potentially false does not mean that everything is up for grabs. There are many relationships in science that have been confirmed so many times that they are termed *laws* because it is extremely doubtful that they will be overturned by future experimentation. It is highly unlikely that we shall find one day that blood does not circulate in the veins and arteries or that the earth does not orbit the sun. These mundane facts are not the type of hypotheses that we have been talking about. They are of no interest to scientists precisely because they are so well established. Scientists are interested only in those aspects of nature that are on the fringes of what is known. They are not interested in things that are so well confirmed that there is no doubt about them.

This aspect of scientific practice—that scientists gravitate to those problems on the fringes of what is known and ignore things that are well confirmed (so-called laws)—is very confusing to the general public. It seems that scientists are always emphasizing what they don't know rather than what is known. This is true, and there is a very good reason for it. To advance knowledge, scientists must be at the outer limits of what is known. Of course, this is precisely where things are uncertain. But science advances by a process of trying to reduce the uncertainty at the limits of knowledge. This can often make scientists look "uncertain" to the public. But this perception is deceiving. Scientists are *only* uncertain at the *fringes* of knowledge—where our understanding is currently being advanced. Scientists are *not* uncertain about the many facts that have been well established by replicable research.

It should also be emphasized that, when scientists talk about falsifying a theory based on observation and about replacing an old, falsified theory with a new one, they do not mean that all the previous facts that established the old theory are thrown out (we shall talk about this at length in Chapter 8). Quite the contrary, the new theory must explain all of the facts that the old theory could explain plus the new facts that the old theory could not explain. So the falsification of a theory does not mean that scientists have to go back to square one. Complex theories can be generally right without being perfectly right, and beliefs can be getting closer to the truth without being exactly true.

Science writer Isaac Asimov illustrated the process of theory revision very well in an essay titled "The Relativity of Wrong" (1989), in which he wrote about how we have refined our notions of the earth's shape. First, he warned us not to think that the ancient belief in a flat earth was stupid. On a plain (where the first civilizations with writing developed), the earth looks pretty flat, and Asimov urged us to consider what a quantitative comparison of different theories would reveal. First, we could express the different theories in terms of how much curvature per mile they hypothesized. The flat-earth theory would say that the curvature is 0 degrees per mile. This theory is wrong, as we know. But in one sense, it is close. As Asimov (1989) wrote,

> About a century after Aristotle, the Greek philosopher Eratosthenes noted that the sun cast a shadow of different lengths at different latitudes (all the shadows would be the same length if the earth's surface were flat). From the difference in shadow length, he calculated the size of the earthly sphere and it turned out to be 25,000 miles circumference. The curvature of such a sphere is about 0.000126 degrees per mile, a quantity very close to 0 per mile, as you can see. . . .
>
> The tiny difference between 0 and 0.000126 accounts for the fact that it took so long to pass from the flat earth to the spherical earth. Mind you, even a tiny difference, such as that between 0 and 0.000126, can be extremely important. The difference mounts up. The earth cannot be mapped over large areas with

any accuracy at all if the difference isn't taken into account and if the earth isn't considered a sphere rather than a flat surface. (pp. 39–40)

But science, of course, did not stop with the theory that the earth was spherical. As we discussed earlier, scientists are always trying to refine their theories as much as possible and to test the limits of current knowledge. For example, Newton's theories of gravitation predicted that the earth should not be perfectly spherical, and indeed this prediction has been confirmed. It turns out that the earth bulges a little at the equator and that it is a little flat at the poles. It is something called an *oblate spheroid.* The diameter of the earth from North Pole to South Pole is 7,900 miles, and the equatorial diameter is 7,927 miles. The curvature of the earth is not constant (as in a perfect sphere); instead, it varies slightly from 7.973 inches to 8.027 inches per mile. As Asimov (1989) noted, "The correction in going from spherical to oblate spheroidal is much smaller than going from flat to spherical. Therefore, although the notion of the earth as a sphere is wrong, strictly speaking, it is not as wrong as the notion of the earth as flat" (p. 41).

Asimov's example of the shape of the earth illustrates for us the context in which scientists use such terms as "*mistake*," "*error*," or "*falsified.*" Such terms do not mean that the theory being tested is wrong in every respect, only that it is *incomplete*. So when scientists emphasize that knowledge is tentative and may be altered by future findings, they are referring to a situation such as this example. When scientists believed that the earth was a sphere, they realized that, in detail, this theory might someday need to be altered. However, the alteration from spherical to oblate spheroidal preserves the "roughly correct" notion that the earth is a sphere. We do not expect to wake up one day and find that it is a cube!

Clinical psychologist Scott Lilienfeld (2005) contextualizes Asimov's point for the psychology student:

> When explaining to students that scientific knowledge is inherently tentative and open to revision, some students may mistakenly conclude that genuine knowledge is impossible. This view, which is popular in certain postmodernist circles, neglects to distinguish knowledge claims that are more certain from those that are less certain. Although absolute certainty is probably unattainable in science, some scientific claims, such as Darwin's theory of natural selection, have been extremely well corroborated, whereas others, such as the theory underpinning astrological horoscopes, have been convincingly refuted. Still others, such as cognitive dissonance theory, are scientifically controversial. Hence, there is a continuum of confidence in scientific claims; some have acquired virtual factual status whereas others have been resoundingly falsified. The fact that methodological skepticism does not yield completely certain answers to scientific questions—and that such answers could in principle be overturned by new evidence—does not imply that knowledge is impossible, only that this knowledge is provisional. (p. 49)

Summary

What scientists most often mean by a *solvable problem* is a *testable theory.* The definition of a testable theory is a very specific one in science: It means that the theory is potentially falsifiable. If a theory is not falsifiable, then it has no implications for actual events in the natural world and, hence, is useless. Psychology has been plagued by unfalsifiable theories, and that is one reason why progress in the discipline has been slow.

Good theories are those that make specific predictions, and such theories are highly falsifiable. The confirmation of a prediction that was not precise. In short, one implication of the falsifiability criterion is that all confirmations of theories are not equal. Theories that receive confirmation from highly falsifiable, highly specific predictions are to be preferred. Even when predictions are not confirmed (i.e., when they are falsified), this falsification is useful to theory development. A falsified prediction indicates that a theory must either be discarded or altered so that it can account for the discrepant data pattern. Thus, it is by theory adjust-

mation of a specific prediction provides more support for the theory from which it was derived than the confirmation

ment caused by falsified predictions that sciences such as psychology get closer to the truth.

Chapter 3
Operationism and Essentialism: "But, Doctor, What Does It Really Mean?"

Learning Objectives

3.1 Explain why science does not answer essentialist questions

3.2 Explain why psychology requires operationism to evaluate theoretical claims

Do physicists really know what gravity is? I mean *really.* What is the real *meaning* of the term gravity? What is the underlying essence of it? What does it ultimately mean even to speak of gravity? When you get down to rock bottom, what is it all about?

Questions such as these reflect a view of science that philosopher Karl Popper called *essentialism.* This is the idea that the only good scientific theories are those that give ultimate explanations of phenomena in terms of their underlying essences or their essential properties. In this chapter, we will discuss why science does not answer essentialist questions such as this and why, instead, science advances by developing *operational definitions* of concepts.

Why Scientists Are Not Essentialists

Scientists, in fact, do not claim to acquire the type of knowledge that the essentialist seeks. The proper answer to the preceding questions is that physicists do *not* know what gravity is in this sense. Science does not attempt to answer "ultimate" questions about the universe. Biologist Peter Medawar (1984) wrote, "There exist questions that science cannot answer and that no conceivable advance of science would empower it to answer. These are the questions that children ask—the 'ultimate questions.' . . . I have in mind such questions as: How did everything begin? What are we all here for? What is the point of living?" (p. 66).

One reason that scientists are suspicious of claims that some person, theory, or belief system provides absolute knowledge about ultimate questions is that scientists consider questions about "ultimates" to be unanswerable. Scientists do not claim to produce perfect knowledge. The unique strength of science is not that it is

an error-free process, but that it provides a way of eliminating the errors that are part of our knowledge base. Furthermore, claims of perfect or absolute knowledge

tend to choke off inquiry. Because a free and open pursuit of knowledge is a prerequisite for scientific activity, scientists are always skeptical of claims that the ultimate answer has been found.

Essentialists Like to Argue About the Meaning of Words

A common indication of the essentialist attitude is an obsessive concern about defining the meaning of terms and concepts before the search for knowledge about them begins. "But we must first define our terms" is a frequent essentialist slogan. "What does that theoretical concept really *mean?*" The idea seems to be that, before a word can be used as a concept in a theory, we must have a complete and unambiguous understanding of all the underlying language problems involved in its usage. In fact, this is exactly the opposite of the way scientists work. Before they begin to investigate the physical world, physicists do not engage in debates about how to use the word *energy* or whether the word *particle* really captures the essence of what we mean when we talk about the fundamental constituents of matter.

The meaning of a concept in science is determined *after* extensive investigation of the phenomena the term relates to, not before such an investigation. The refinement of conceptual terms comes from the interplay of data and theory that is inherent in the scientific process, not from debates on language usage. Essentialism leads us into endless argument about words. Instead, scientists believe that such language games distract us from matters of substance. For example, concerning the question "What is the true meaning of the word *life*?" two biologists give the startling answer that "There is no true meaning. There is a usage that serves the purposes of working biologists well enough, and it is not the subject of altercation or dispute" (Medawar & Medawar, 1983, pp. 66–67). In short, the explanation of phenomena, not the analysis of language, is the goal of the scientist. The key to progress in all the sciences has been to abandon essentialism and to adopt operationism, our topic of inquiry in this chapter.

Operationists Link Concepts to Observable Events

Where, then, does the meaning of concepts in science come from if not from discussions about language? What are the criteria for the appropriate use of a scientific concept? To answer these questions, we must discuss operationism, an idea that is crucial to the construction of theory in science and one that is especially important for evaluating theoretical claims in psychology.

Although there are different forms of operationism, it is most useful for the consumer of scientific information to think of it in the most general way. *Operationism* is simply the idea that concepts in scientific theories must in some way be grounded in, or linked to, observable events that can be measured. Linking the concept to an observable event makes the concept public. The operational definition removes the concept from the feelings and intuitions of a particular individual and allows it to be tested by anyone who can carry out the measurable operations.

For example, defining the concept *hunger* as "that gnawing feeling I get in my stomach" is not an operational definition because it is related to the personal experience of a "gnawing feeling" and, thus, is not accessible to other observers. By contrast, definitions that involve some measurable period of food deprivation or some physiological index such as blood sugar levels are operational because they involve observable measurements that anyone can carry out. Similarly, psychologists cannot be content with a definition of *anxiety*, for example, as "that uncomfortable, tense feeling I get at times" but must define the concept by a number of operations such as questionnaires and physiological

define the concept by a number of operations such as questionnaires and physiological measurements. The former definition is tied to a personal interpretation of bodily states and is not replicable by others. The latter puts the concept in the public realm of science.

It is important to realize that a concept in science is defined by a *set* of operations, not by just a single behavioral event or task. Instead, several slightly different tasks and behavioral events are used to converge on a concept (we will talk more about the idea of converging operations in Chapter 8). For example, educational psychologists define a concept such as *reading ability* in terms of performance on a standardized instrument such as the Woodcock Reading Mastery Tests (Woodcock, 2011) that contains a whole *set* of tasks. The total reading ability score on the Woodcock Reading Mastery instrument comprises indicators of performance on a number of different subtests that test slightly different skills, for example, reading a passage and thinking of an appropriate word to fill in a blank in the passage, coming up with a synonym for a word, pronouncing a difficult word correctly in isolation, and several others. Collectively, performance on all of these tasks defines the concept *reading ability*.

Operational definitions force us to think carefully and empirically—in terms of observations in the real world—about how we want to define a concept. Imagine trying to define operationally something as seemingly conceptually simple as typing ability. Imagine you need to do this because you want to compare two different methods of teaching typing. Think of all the decisions you would have to make. You would want to measure typing speed, of course. But over how long a passage? A passage of only 100 words would seem too short, and a passage of 10,000 words would seem too long. But exactly how long then? How long does speed have to be sustained to match how we best conceive of the theoretical construct *typing ability*? And what kind of material has to be typed? Should it include numbers and formulas and odd spacing? And how are we going to deal with errors? It seems that both time and errors should come into play when measuring typing ability, but exactly what should the formula be that brings the two together? Do we want speed and errors to be equally weighted, or is one somewhat more important than the other? The need for an operational definition would force you to think carefully about all of these things; it would make you think very thoroughly about how to conceptualize typing ability.

Learn to develop the habit of inquiring about the details of operational definitions. Cognitive scientist Dan Levitin (2016) gives us the example of an advocacy group claiming that: 70 percent of schoolchildren ages ten to eighteen are sexually active. First, ages ten to eighteen is a huge range of maturity. Also, with a percentage so high, mathematically it must be the case that there is a substantial percentage at ages as low as ten and eleven who are sexually active. This adds to our curiosity about how the concept "sexual activity" is being operationally defined. To know how to interpret this high percentage, we would need to really drill down on *precisely* how the group promoting this statistic is defining the concept "sexual activity"—and this we will not do because this is not an X-rated book!

Popular author Michael Lewis (2017), in his book *The Undoing Project*, describes how, years ago, the Houston Rockets of the National Basketball Association tried to develop better methods of measuring a player's performance than had been used before. For example, rather than just measuring the number of rebounds a player got, they measured the number of rebounds divided by the number of *opportunities* for rebounds (that is, they measured the *proportion* of successful rebounds). They moved all of their statistics (points, steals, rebounds, etc.) from a per-game basis to a per-minute played basis. What the Rockets were essentially trying to do was to develop *better*, more refined operational definitions of player impact than had been used before. Their efforts demonstrate that better operational definitions are often more specific.

Finally, consider the task of the Food and Drug Administration, which has to decide what is an "unacceptable" level of contamination for various foods as opposed to what are considered "unavoidable defects" (Levy, 2009). A federal

agency such as the FDA cannot be subjective about such things. It needs strict operational definitions of its judgments with respect to contaminants in each food that it inspects. So, for example, it comes up with operational definitions of the following

sort (Levy, 2009): An "unacceptable" level of contamination in tomato juice is more than 10 fly eggs per 100 grams; an "unacceptable" level of contamination in mushrooms is five or more maggots 2 millimeters or longer per 100 grams. Very gross—but commendably operational!

Reliability and Validity

Operationalizing a concept in science involves *measurement:* assigning a number to an observation via some rule. Science writer Charles Seife (2010) makes the point that once we start using numbers in measurement we suddenly start caring about them. His argument is that the nonmathematician rarely cares about the properties of numbers when they are used merely as abstract symbols. We don't care about the number five, by itself. But as soon as the number five becomes five "pounds" or five "dollars" or five "percent inflation" or five "IQ points"—then suddenly we start to care. Seife (2010) says that "a number without a unit is ethereal and abstract. With a unit, it acquires meaning—but at the same time, it loses its purity" (p. 9). What Seife means by "losing its purity" is that once we are involved with measurement—once the number has a unit attached—we are suddenly concerned that numbers have the "right" properties. What are the "right" properties for a number to have in science? The answer to this question is that, in science, the "right" properties for a number to have are the properties of reliability and validity.

For an operational definition of a concept to be useful, it must display both reliability and validity. *Reliability* refers to the consistency of a measuring instrument—whether you would arrive at the same measurement if you assessed the same concept multiple times. The scientific concept of reliability is easy to understand because it is very similar to its layperson's definition and very like one of its dictionary definitions: "an attribute of any system that consistently produces the same results."

Consider how a layperson might talk about whether something was reliable or not. Imagine a New Jersey commuter catching the bus to work in Manhattan each morning. The bus is scheduled to arrive at the commuter's stop at 7:20 A.M. One week the bus arrives at 7:20, 7:21, 7:20, 7:19, and 7:20, respectively. We would say that the bus was pretty reliable that week. If the next week the bus arrived at 7:35, 7:10, 7:45, 7:55, and 7:05, respectively, we would say that the bus was very unreliable that week.

The reliability of an operational definition in science is assessed in much the same way. If the measure of a concept yields similar numbers for multiple measurements of the same concept, we say that the measuring device displays high reliability. If we measured the same person's intelligence with different forms of the same IQ test on Monday, Wednesday, and Friday of the same week and got scores of 110, 109, and 110, we would say that that particular IQ test seems to be very reliable. By contrast, if the three scores were 89, 130, and 105, we would say that that particular IQ test does not seem to display high reliability. There are specific statistical techniques for assessing the reliability of different types of measuring instruments, and these are discussed in all standard introductory methodology textbooks.

But remember that reliability is only about consistency and nothing else. Reliability alone is not enough for an operational definition to be adequate. Reliability is necessary but not sufficient. A good operational definition of a concept must also be a *valid* measure of that concept. The term *construct validity* refers to whether a measuring instrument (operational definition) is measuring what it is supposed to be measuring. In his methodology textbook, psychologist Paul Cozby (2014) gives us a humorous example of reliability without validity. Imagine you are about to get your

intelligence assessed. The examiner tells you to stick out your foot and clamps on a measuring device like those at the shoe store and reads out a number. You would, of course, think that this was a joke. But note that this measuring instrument would

display many of the types of reliability that are discussed in methodology textbooks. It would give virtually the same readings on Monday, Wednesday, and Friday (what is termed *test–retest reliability*) and it would give the same reading no matter who used it (what is termed *interrater reliability*).

The problem with the shoe device as a measure of intelligence is not reliability (which it has) but validity. It is not a good measure of the concept it purports to measure (intelligence). One way we would know that it is not a valid measure of intelligence is that we would find that it does not relate to many other variables that we would expect a measure of intelligence to relate to. Measures from the shoe instrument would not relate to academic success; they would not relate to neurophysiological measures of brain functioning; they would not relate to job success; and they would not relate to measures of the efficiency of information processing developed by cognitive psychologists. By contrast, actual measures of intelligence relate to all of these things (Deary, 2013; Duncan, 2010; Hunt, 2011; Ritchie, 2015; Sternberg & Kaufman, 2011). Actual measures of intelligence in psychology have validity as well as reliability, whereas the shoe-size measure of intelligence has reliability without validity.

You might be wondering about another combination of affairs at this point, so let me recapitulate where we are. In operational definitions, we are looking for both reliability and validity, so high reliability and high validity are sought. We have just discussed the shoe-size IQ test in order to demonstrate that high reliability and low validity get us nowhere. A third case, low reliability and low validity, is so obviously useless that it is not worth discussing. But you might be wondering about the fourth and last possible combination: What if something has high validity and low reliability? The answer is that, like its converse case of low validity and high reliability (the shoe-size example), this state of affairs gets you nowhere. And, actually, it is more accurate to say that this state of affairs is impossible—because you cannot claim to be measuring validly if you cannot measure reliably.

When trying to develop valid operational definitions it is important to be precise about just what concept we are trying to measure. For example, the National Football League evaluates quarterbacks using a construct that it calls the “passer rating” (Sielski, 2010). It is important to realize that this construct is named precisely as “passer rating.” That is, it is specifically *not* a quarterback rating. This is because the operational definition of “passer rating” takes into account only passing and not everything a quarterback does. Specifically, the passer rating is a mathematical formula that involves the following four things: pass completion percentage, yards per passing attempt, touchdowns per passing attempt, and interceptions per attempt. The passer rating statistic does not involve: the running yards gained by the quarterback, play-calling ability, win-loss record, sacks, fumbles, and a variety of other quantifiable quarterback variables. For this reason, another construct with a different operational definition has been developed called the “total quarterback rating.”

Ultimately, what measurement is about is interpreting numbers in a meaningful context. A simple number, without its proper context, is meaningless, or worse, it is misleading. Mathematics professor Jordan Ellenberg (2015) tells us of a blogger who warned about the air pumped into airplane cabins because it was mixed with high levels of nitrogen (“sometimes almost as high as 50%”). The blogger failed to mention that the natural proportion of nitrogen in the Earth’s atmosphere is 78 percent! Likewise, many people are surprised to learn that the number of bank tellers in the United States is the same as it was in 1980. They are surprised that automation hasn’t destroyed a lot of the teller jobs. But automation *has* destroyed many such jobs. The statement presented (“same number as in 1980”) fails to take into account the proper

context. The proper context would take into account the fact that the population has increased by 40 percent since 1980. The *proportion* of teller jobs has gone down because, indeed, many such jobs have been lost to technology.

Interpreting numbers out of their proper context can have profound practical consequences. Science writer Gina Kolata (2016a) describes how many more men with ambiguous prostate cancer test results are now choosing active surveillance rather than radiation or surgery (up to 40 percent, from just 10 percent a couple of decades ago). This is because it has been found that men with a Gleason score of 6 or less have a less than 1 percent chance of dying from prostate cancer in the next 10 years. The Gleason score is a pathologist's assessment of how ominous the prostate cells look, and it starts at 2 and goes to 10. Despite the fact that it sounds like a high number, a score of 6 is actually the lowest score for cells that can be called cancer. It sounds high in a 2 to 10 context, and it scares a lot of men into having surgery. For this reason, the World Health Organization has renamed a Gleason 6 score a score of Group 1 in a scale that goes from 1 to 5 (Kolata, 2016a). This context reframes (Stanovich et al., 2016; Thaler, 2015) the score more accurately as one that indicates cancer but at an early stage.

Direct and Indirect Operational Definitions

The link between concepts and observable operations varies greatly in its degree of directness or indirectness. Few scientific concepts are defined almost entirely by observable operations in the real world. Most concepts are defined more indirectly. For example, the use of some concepts is determined by both a set of operations and the particular concept's relationship to other theoretical constructs. Finally, there are concepts that are not directly defined by observable operations but are linked to other concepts that are. These are sometimes called latent constructs, and they are common in psychology.

For example, much research has been done on the so-called type A behavior pattern because it has been linked to the incidence of coronary heart disease (Boehm & Kubzansky, 2012; Chida & Hamer, 2008; Martin et al., 2011; Matthews, 2013; Mostofsky et al., 2014). We will discuss the type A behavior pattern in more detail in Chapter 8. The important point to illustrate here, however, is that the type A behavior pattern is actually defined by a set of subordinate concepts: a strong desire to compete, a potential for hostility, time-urgent behavior, an intense drive to accomplish goals, anger, and several others. However, each one of these defining features of the type A behavior pattern (a strong desire to compete, hostility, etc.) is *itself* a concept in need of operational definition. Indeed, considerable effort has been expended in operationally defining each one. The important point for our present discussion is that the concept of the type A behavior pattern is a complex concept that is not directly defined by operations. Instead, it is linked with other concepts, which, in turn, have operational definitions.

The type A behavior pattern provides an example of a concept with an indirect operational definition. A similarly defined concept in clinical psychology is that of distress tolerance (Zvolensky et al., 2010). This global concept is defined in terms of several simpler subconstructs more closely tied to operational measures: tolerance of uncertainty, tolerance of ambiguity, tolerance of frustration, tolerance of negative emotion, and tolerance of physical discomfort.

In short, although theoretical concepts differ in how closely they are linked to observations, all concepts acquire their meaning partially through their link to such observations.

Scientific Concepts Evolve

It is important to realize that the definition of a scientific concept is not fixed but constantly changing as the observations that apply to the concept are enriched. If the original operational definition of a concept turns out to be theoretically unfruitful, it will be abandoned in favor of an alternative set of defining operations. Thus, concepts in science are continually evolving and can increase in abstractness as the knowledge

concerning them increases. For example, at one time the electron was thought of as a tiny ball of negative charge circling the nucleus of an atom. Now it is viewed as a probability density function having wavelike properties in certain experimental situations.

In psychology, the development of the concept of intelligence provides a similar example. At first, the concept had only a strict operational definition: Intelligence is what is measured by tests of mental functioning. As empirical evidence accumulated relating intelligence to scholastic achievement, learning, brain injury, neurophysiology, and other behavioral and biological variables, the concept was both enriched and refined (Deary, 2013; Deary et al., 2010; Duncan, 2010; Plomin et al., 2016; Shipstead et al., 2016). It now appears that intelligence is best conceptualized as a higher order construct defined by several more specific information-processing operations. These hypothesized processes, in turn, have more direct operational definitions stated in terms of measurable performance.

The concepts in theories of human memory have likewise evolved. Psychologists now rarely use global concepts like *remembering* or *forgetting;* instead, they test the properties of more specifically defined memory subprocesses, such as short-term acoustic memory, iconic storage, semantic memory, and episodic memory. The older concepts of remembering and forgetting have been elaborated with more specifically operationalized concepts. Recall our earlier discussion of how the Houston Rockets developed measures of player effectiveness that were increasingly refined.

As scientific concepts evolve, they often become enmeshed in several different theoretical systems and acquire alternative operational definitions. There is not necessarily anything wrong with the concept when this happens. For example, you would think that the measurement of something like the unemployment rate would be stable and uncomplicated. But if you thought that, you would be wrong. The standard number that you hear on the radio, or read about in the newspaper, is just one of six different ways that economists measure the unemployment rate (Bureau of Labor Statistics, 2014, 2017; Craven McGinty, 2016). The statistic you usually hear on television is known as U3 and it categorizes people as unemployed if they do not have a job, have actively looked for work in the prior four weeks, and are currently available for work. Another measure, known as U1 calls people unemployed only if they have been out of work for 15 weeks or longer. Because this definition is more restrictive than U3, it leads to unemployment rates that are lower than those indicated by U3. U6 is at the opposite extreme from U1. U6 adds to U3 the following: discouraged workers, part-time workers who want full-time jobs, and marginally attached workers. Thus, U6 always leads to a higher number than U3 and to a much higher number than U1. There are three others that are slightly different from U1, U3, and U6 and that lead to somewhat different unemployment rates. They are all alternative operational definitions of the concept.

Despite examples like the unemployment case, where other disciplines have multiple alternative operational definitions of their concepts, many people still believe that psychology is discredited by the fact that many of its important theoretical constructs are operationalized and conceptualized in more than one way. But such a situation is not unique to psychology, and it is not a matter for despair or hand-wringing. In fact, it is a relatively common occurrence in science. Heat, for example, is conceptualized in terms of thermodynamic theory and in terms of kinetic theory. Physics is not scandalized by this state of affairs. Likewise, consider the electron. Many of its properties are explained by its being conceptualized as a wave. Other properties, however, are better handled if it is viewed as a particle. The existence of these alternative conceptualizations has tempted no one to suggest that physics be abandoned.

People got a lesson in this point in 2006 when the media reported that the International Astronomical Union had recently reoperationalized the term "planet" in a way that excluded Pluto (Adler, 2006; Brown, 2010). That something as seemingly basic as the concept "planet" could be the subject of alternative views was a surprise

to many in the public. But in fact it is a common occurrence in science. In this case, one group of astronomers prefers to stress the geologic composition of astronomical bodies. Another group likes to emphasize their dynamical properties, for example, their orbits and gravitational effects. The operational definitions of the former group

include Pluto as a planet, but the operational definitions of the latter group exclude Pluto (Layton & Koh, 2015). Three things are needed, according to this operational definition: a body must orbit the sun; it must be roughly spherical; and it must “clear its neighborhood” (which means dominating its orbit by pushing away smaller bodies or turning them into satellites). Pluto does not fulfill this last condition (although it does fulfill the other two). Thus, it was demoted to the status of “dwarf planet.” These disputed, and differing, operational definitions do not reflect badly on the field of astronomy. They merely reflect different ways of triangulating concepts in the discipline. The same is true in psychology, where there are sometimes alternative operational definitions of concepts. Just because something is hard to define does not mean that there is not something real there to study.

Operational Definitions in Psychology

Many people understand the necessity of operationism when they think about physics or chemistry. They understand that if scientists are going to talk about a particular type of chemical reaction, or about energy, or about magnetism, they must have a way of measuring these things. Unfortunately, when people think and talk about psychology, they often fail to recognize the need for operationism. Why is it not equally obvious that psychological terms must be operationally defined, either directly or indirectly, in order to be useful explanatory constructs in scientific theories?

One reason is what has been termed the *preexisting-bias problem* in psychology. The problem is this: People do not come to the study of geology with emotionally held beliefs about the nature of rocks. The situation in psychology is very different. Psychology professor Elizabeth Phelps (2013) of New York University has joked that her mother, who is a marine biologist, never has non-scientist friends telling her that her studies of pollutants in estuaries do not “make sense,” but that everyone seems to have an opinion on Phelps’ own research into memory and decision biases.

We all have intuitive theories of personality and human behavior because we have been “explaining” behavior to ourselves all our lives. All our personal psychological theories contain theoretical concepts (e.g., *smart, aggressive, anxiety*). One of the greatest sources of misunderstanding and one of the biggest impediments to the accurate presentation of psychological findings in the media is the fact that many technical concepts in psychology are designated by words used in everyday language. This everyday usage opens the door to a wide range of misconceptions. The layperson seldom realizes that when psychologists use words such as *intelligence, anxiety, aggression,* and *attachment* as theoretical constructs, they do not necessarily mean the same thing that the general public does when using these words.

The nature of this difference should be apparent from the previous discussion of operationism. When terms such as *intelligence* and *anxiety* are used in psychological theories, their direct or indirect operational definitions determine their correct usage. These definitions are often highly technical, usually fairly specific, and often different from popular usage in many ways. For example, when hearing the phrase “the first principal component of the factor analysis of a large sampling of cognitive tasks,” many people would not recognize it as part of the operational definition of the term *intelligence.*

Similarly, in lay usage, the word *depression* has come to mean something like “feeling down in the dumps.” By contrast, the technical definition of major depressive disorder takes up over a dozen pages in the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000) and means something quite different from being “down in the dumps.” A clinical psychologist’s depression is not

the same as the layperson's depression (Klein, 2010). Such a basic and simple word as "significant" means something different to a psychologist than it does to a layperson (Levitin, 2016). To the latter it means something like noteworthy, or important.

In psychological discourse about research, it means something much more specific and technical: It means *statistical* significance—an assessment of how likely the results would have been obtained purely on the basis of chance.

When the psychologist and the layperson use the same word to mean different things, they often misinterpret each other. Such confusion would be less prevalent if new words had been coined to represent psychological constructs. On occasion such words have been coined. Just as physicists have their erg and joule, psychology has its dissonance and encoding, words that are not actually coined but are uncommon enough to prevent confusion.

"But," the layperson may object, "why do psychologists inflict this on us? New jargon, highly technical definitions, uncommon uses of words. Why do we need them? Why is my idea of 'intelligence' not an acceptable idea to talk about?"

Here we see exemplified a critical misunderstanding of psychological research—a misunderstanding that is often reflected in media reports of psychological research. A national newspaper report on the 1996 meeting of the American Psychological Association (Immen, 1996) is headlined "Could You Repeat That in Klingon?" and refers to "psychologists speaking a language all their own." The article ridicules the following title of a paper delivered at the conference: "Interpreting WJ-R and KAIT Joint Factor Analyses from Gf-Gc Theory." Although the reporter states that he would "not even dare to speculate about the true meaning" of the title, almost all properly trained psychologists would recognize the title as referring to developments in intelligence test theory. And this is as it should be. Gf-Gc theory is a technical development in intelligence theory. There is no reason for the reporter to have heard of this concept—just as one would not expect the reporter to know the details of the latest elementary particle to be identified by physicists. Somehow, however, the reporter's (quite understandable) ignorance of the technical terminology is seen as reflecting negatively on modern psychology. When the topic is physics, reporters seem to know that it is their own ignorance that is preventing understanding. But when the topic is psychological, they act as if psychologists are to blame for their own lack of understanding.

We come here to the crux of the problem. The first step in resolving it is to emphasize a point from our earlier discussion: Operationism is not unique to psychology. It is characteristic of all sciences. Most of the time, we accept it readily, recognizing its obvious nature. If a scientist is investigating radioactivity, we take it for granted that he or she must have some observable way of measuring the phenomenon—a method that another investigator could use to obtain the same results. This method is what makes possible the public nature of science, one of its defining features. Two different scientists agree on the same operational definition so that it is possible for one to replicate the other's results. However, what seems obvious in other contexts is sometimes not so clear when we think about psychology. The necessity for operational definitions of concepts like *intelligence* and *anxiety* is often not recognized because we use these terms all the time, and, after all, don't we all just "know" what these things mean?

The answer is "No, we don't"—not in the sense that a scientist has to know—that is, in a public sense. A scientist must "know" what intelligence means by being able to define, precisely, how another laboratory could measure it in exactly the same way and be led to the same conclusions about the concept. This is vastly different—in terms of explicitness and precision—than the vague verbal connotations that are needed in order to achieve casual understanding in general conversation.

Operationism as a Humanizing Force

The problem with relying on what we all just "know" is the same problem that plagues all intuitive (i.e., nonempirical) systems of belief. What you "know" about something may not be quite the same as what Jim "knows" or what Jane "knows." How do we

decide who is right? You may say, "Well, I feel strongly about this, so strongly that I *know* I'm right." But what if Jim, who thinks somewhat differently, feels even more strongly than you do? And then there's Jane, who thinks differently from you or Jim, claiming that she must be right because she feels *even more* strongly than Jim does.

This simple parody is meant only to illustrate a fundamental aspect of scientific knowledge, one that has been a major humanizing force in human history: In science, the truth of a knowledge claim is not determined by the strength of belief of the individual putting forth the claim. The problem with all intuitively based systems of belief is that they have no mechanism for deciding among conflicting claims. When everyone knows intuitively, but the intuitive claims conflict, how do we decide who is right? Sadly, history shows that the result of such conflicts is usually a power struggle.

Some people mistakenly claim that an operational approach to psychology dehumanizes people and that instead we should base our views of human beings on intuition. Instead, the truly humane position is one that bases theoretical views of human beings on observable behavior rather than on the feelings of the theorizer. Science makes knowledge claims public so that conflicting ideas can be tested in a way that is acceptable to all disputants. Science puts observation in place of a power struggle. This allows a selection among theories to take place by peaceful mechanisms that we all agree on in advance. The public nature of science rests critically on the idea of operationism. By operationally defining concepts, we put them in the public realm, where they can be criticized, tested, improved, or perhaps rejected.

Psychological concepts cannot rest on someone's personal definition, which may be uncommon, idiosyncratic, or vague. For this reason, psychology must reject all personal definitions of concepts—just as physics, for example, rejects personal definitions of energy and meteorology rejects personal definitions of what a cloud is. Psychologists instead must rely on publicly accessible concepts defined by operations that anyone with proper training and facilities can perform. In rejecting personal definitions, psychology is not shutting out the layperson but is opening up the field—as all sciences do—to the quest for a common, publicly accessible knowledge that all can share.

Essentialist Questions and the Misunderstanding of Psychology

Another reason many people seem to abandon the idea of operationism when they approach psychology is that they seek essentialist answers to certain human problems. Recall the questions at the beginning of this chapter: What is the real meaning of the word *gravity*? What is the underlying essence of it? What does it ultimately mean even to speak of gravity? Most people would recognize that these questions require knowledge of the ultimate, underlying nature of a phenomenon and that current theories in physics cannot provide answers to questions of this type. Anyone familiar with popular writing about the progress of physical science in the last few centuries will recognize that gravity is a theoretical construct of great complexity and that its conceptual and operational relationships have been in constant flux.

However, substitute the word intelligence for the word gravity in each of the preceding questions ("what is the essence of intelligence," "what does it really mean to say that someone is intelligent") and, suddenly, a miracle occurs. Now the questions are imbued with great meaning. They seem natural and meaningful. They literally beg for an ultimate answer. When the psychologist gives the same answer as the physicist—that intelligence is a complex concept that derives meaning from the operations used to measure it and from its theoretical relationships to other constructs—he or she

is often belittled and accused of avoiding the real issues.

One problem facing psychology, then, is that the public demands answers to essentialist questions that it does not routinely demand of other sciences. These demands often underlie many of the attempts to disparage the progress that has been made in the

field. Although these demands do not hinder the field itself—because psychologists, like other scientists, ignore demands for essentialist answers and simply go about their work—they are an obstacle to the public's understanding of psychology. The public becomes confused when an uninformed critic claims that there has been no progress in psychology. The fact that this claim so frequently goes unchallenged reflects the unfortunate truth of the major premise of this book: Public knowledge of what scientific achievement within psychology actually means is distressingly meager. When examined closely, such criticisms usually boil down to the contention that psychology has not yet provided the ultimate answer to any of its questions. To this charge, psychology readily pleads guilty—as do all the other sciences.

Some may find it discomforting to learn that no science, including psychology, can give answers to essentialist questions. Think of physics and the phenomenon of radioactive decay in which the number of atoms of a radioactive element that have decayed can be related to time via an exponential mathematical function. That function, however, does not explain *why* radioactive decay occurs. It does not answer the layperson's question of *why* it follows this function. It does not answer the question of what radioactive decay *really is.* Similarly, physics makes no attempt to explain why things obey laws of electromagnetism or of gravitation. Physics does not explain what things *really* are or *why* they got that way.

Likewise, those who seek essentialist answers to questions concerning human nature are destined to be disappointed if they are looking to psychology. Psychology is not a religion. It is a broad field that seeks a scientific understanding of all aspects of behavior. Therefore, psychology's current explanations are temporary theoretical constructs that account for behavior better than alternative explanations. These constructs will certainly be superseded in the future by superior theoretical conceptualizations that are closer to the truth.

The idea of an operational definition can be a very useful tool in evaluating the falsifiability of a psychological theory. The presence of concepts that are not directly or indirectly grounded in observable operations is an important clue to recognizing a nonfalsifiable theory. Thus, the presence of loose concepts—those for which the theorist cannot provide direct or indirect operational links—should be viewed with suspicion.

A principle that scientists term *parsimony* is relevant here. The principle of parsimony dictates that when two theories have the same explanatory power, the simpler theory (the one involving fewer concepts and conceptual relationships) is preferred (Gallistel, 2016). The reason is that the theory with fewer conceptual relationships will likely be the more falsifiable of the two in future tests.

Summary

Operational definitions are definitions of concepts stated in terms of observable operations that can be measured. One of the main ways that we ensure that theories are falsifiable is by making certain that the key concepts in theories have operational definitions stated in terms of well-replicated behavioral observations. Operationally defined concepts make scientific knowledge publicly verifiable. Such definitions are in the public domain so that the theoretical concepts that they define are testable by all—unlike "intuitive," people have preexisting notions about what these terms mean, the necessity of operationally defining these terms is often not recognized. Psychology is like all other sciences in requiring operational definitions of its terms. However, people often demand answers to essentialist questions (questions about the absolute, underlying nature of a concept) of psychology that they do not demand of other sciences. No science provides such answers to ultimate questions. Instead, psychology, like other sciences, seeks

cepts that they define are testable by all—unlike intuitive, nonempirical definitions that are the special possession of particular individuals and not open to testing by everyone.

Because psychology employs terms from common discourse, such as *intelligence* and *anxiety*, and because many

questions. Instead, psychology, like other sciences, seeks continually to refine its operational definitions so that the concepts in theories more accurately reflect the way the world actually is.

Chapter 4

Testimonials and Case Study Evidence: Placebo Effects and the Amazing Randi

Learning Objectives

4.1 Outline the limitations of case studies and testimonials as scientific support for theories

4.2 Explain how the placebo effect invalidates testimonial or case study evidence

4.3 Explain how vividness of information affects interpretation of scientific evidence

4.4 Describe how to identify pseudoscientific claims

Television talk shows that presented a lot of self-help advice were quite popular in the 1990s and 2000s (The Oprah Winfrey Show was the top-rated of these programs). The show host would often present a so-called "expert" guest who would often answer audience questions. Here's an example of how it might go.

Today's guest is Dr. Alfred Pontificate, director of the Oedipus Institute of Human Potential. The host attempts to elicit questions about the doctor's provocative new Theory of Birth Order, which is based on the idea that the course of one's life is irrevocably set by family interactions that are determined by birth order. The discussion inevitably turns from theoretical concerns to requests for explanations of personal events of importance to members of the audience. The doctor complies without much prodding.

For example, "Doctor, my brother is a self-destructive workaholic. He ignores his wife and family and places work-related problems above everything else. He has an ulcer and a drinking problem that he refuses to acknowledge. His family hasn't been on a real vacation in two years. He's headed for divorce and doesn't seem to care. Why has he chosen such a self-destructive course?"

To which the doctor replies, "What is his birth order, my dear?"

"Oh, he is the oldest of the children."

"Yes," the doctor says, "this is quite common. We see it often in the clinic. The

underlying dynamics of a situation like this arise because parents transfer their life hopes and frustrations to their firstborn child. Through a process of unconscious wish transference, the child absorbs these hopes and frustrations, even if the parents never articulate them. Then, through the unconscious process that I call the dynamic expectation spiral,

the aspirations of the parents become manifest as a pathological need for achievement in the child."

Although the audience members on the show sometimes ask hostile questions when the guest challenges their beliefs, this rarely happens when a behavioral "expert" seems to confirm conventional wisdom. Once in a while, however, the show is enlivened by an audience member who questions the evidence behind the guest's declarations. In this case, an eager, forthright questioner is in the studio. "But wait a minute, Doctor," the questioner begins. "My brother is a firstborn, too. My parents sent the bum to Harvard and told me to go to a two-year school to be a dental hygienist. So, this 'great brain' of theirs drops out after one year, goes to some mountaintop in Colorado, and the last time we saw him he was weaving baskets! I don't understand what you're saying about firstborns."

The audience tenses for the confrontation, but alas, the doctor always wins in the end: "Oh, yes, I have seen many cases like your brother. Yes, I often meet them in my practice. They are people for whom the dynamic expectation spiral has short-circuited, creating an unconscious desire to thwart wish transference. Thus, the individual's life develops in such a way as to reject conventional achievement aspirations." A hushed pause follows; then on we go to the next "case."

Of course, we are dealing with something quite familiar here. This is another example of the Benjamin Rush problem discussed in Chapter 2. This "theory" of birth order is structured so that no observation can disconfirm it. Because it is an unfalsifiable theory, the confirmations put forth to prove it are meaningless because nothing is ruled out by the theory.

However, our concern in this chapter is not with the theory itself, but with the nature of the evidence that is presented to support it. When pressed for evidence, Dr. Pontificate presents his own "clinical experience" or "case studies" as proof. This is an extremely common occurrence in the realm of media psychology. Talk shows, websites, and paperback book racks are full of psychological theories based on the clinical experience of the author. Many of the therapies presented to the public through these outlets are backed by nothing more than the testimonials of individuals who have undergone them and consider themselves improved or cured. In this chapter, we shall develop a principle of great use to consumers of psychological information: Case studies and testimonials are virtually worthless as evidence for the evaluation of psychological theories and treatments.

In this chapter, we will demonstrate why this is true, and we will also discuss the proper role of the case study in psychology.

The Place of the Case Study

A case study is an investigation that looks intensely and in detail at a single individual or very small number of individuals. The usefulness of case study information is strongly determined by how far scientific investigation has advanced in a particular area. The insights gained from case studies or clinical experience may be quite useful in the early stages of the investigation of certain problems as indicators of which variables deserve more intense study. Case studies have played a prominent role in opening up new areas of study in psychology (Martin & Hull, 2006). Well-known examples occur in the work of Jean Piaget. Piaget's investigations raised the possibility that children's thinking is not just a watered-down or degraded version of adults' thinking but has a structure of its own. Some of Piaget's conjectures about children's thinking have been confirmed, but

many have not (Bjorklund & Causey, 2017; Goswami, 2013). However, what is important for our discussion here is not how many of Piaget's conjectures have been confirmed.

Instead, what is important is to understand the fact that Piaget's case studies did not *prove* anything but, rather, suggested incredibly fruitful areas for developmental psychologists to investigate. It was subsequent correlational and experimental studies of the type to be described in Chapters 5 and 6 that provided the confirmatory and disconfirmatory evidence for the hypotheses that were generated by Piaget's case studies.

When we move from the early stages of scientific investigation, where case studies may be very useful, to the more mature stages of theory testing, the situation changes drastically. Case studies are not useful at the later stages of scientific investigation because they cannot be used as confirming or disconfirming evidence in the test of a particular theory. The reason is that case studies and testimonials are isolated events that lack the comparative information necessary to rule out alternative explanations.

One of the limitations of Freud's work was that he never took the second step of moving from interesting hypotheses based on case studies to actually *testing* those hypotheses (Boudry & Buekens, 2011). One of the leading writers about Freud's work, Frank Sulloway, has said that "science is a two-step process. The first step is the development of hypotheses. Freud had developed a set of extremely compelling, extremely plausible hypotheses for his day, but he never took that key, second procedural step in the rigorous manner that is required for true science" (Dufresne, 2007, p. 53).

Testimonials are like case studies in that they are isolated events. The problem of relying on testimonial evidence is that there are testimonials to support virtually every therapy tried. Thus, it is wrong to use them to support any *specific* remedy, because all of the competing remedies *also* have supporting testimonials. What we want to know, of course, is which remedy is *best*, and we cannot determine this by using testimonial evidence. That is because every pseudoscientific medical cure has produced sincere patients who will give testimonials that the remedy "really worked for them!" For example, subliminal self-help audiotapes (tapes that use messages below hearing threshold) that are purported to raise memory performance or self-esteem generate plenty of testimonials despite the fact that controlled studies indicate that they have absolutely no effect on memory or self-esteem (Lilienfeld et al., 2010).

The idea of alternative explanations is critical to an understanding of theory testing. The goal of experimental design is to structure events so that support of one particular explanation simultaneously disconfirms other explanations. Scientific progress can occur only if the data that are collected rule out some explanations, as discussed in Chapter 2 on falsifiability. Science sets up conditions for the natural selection of ideas. Some survive empirical testing and others do not. Those that remain are closer to the truth. This is the honing process by which ideas are sifted so that those that contain the most truth are found. But there must be selection and elimination in this process: Data collected as support for a particular theory must not leave many other alternative explanations as equally viable candidates. For this reason, scientists construct control or comparison groups in their experimentation. Control groups are formed so that, when their results are compared with those from an experimental group, some alternative explanations are ruled out. How this is done will be a main topic in several later chapters.

Case studies and testimonials stand as isolated phenomena. They lack the comparative information necessary to prove that a particular theory or therapy is superior. It is thus wrong to cite a testimonial or a case study as support for a *particular* theory or therapy. Those who do so mislead the public if they do not point out that such evidence is open to a wide range of alternative explanations. In short, the isolated demonstration of a phenomenon may be highly misleading. This point can be illustrated more specifically by the example of placebo effects.

Why Testimonials Are Worthless: Placebo Effects

Virtually every therapy that has ever been devised in medicine and psychology has garnered supporters and has been able to produce individuals who will testify sincerely to its efficacy. Medical science has documented testimonials to the curative powers of swine teeth, crocodile dung, powdered Egyptian mummy, and many other even more imaginative

remedies (Harrington, 2008). In fact, it has long been known that the mere suggestion that treatment is being administered is enough to make many people feel better.

The tendency of people to report that any treatment has helped them, regardless of whether it has a real therapeutic element, is known as the *placebo effect* (Benedetti et al., 2011; Churchland, 2015; Lu, 2015; Marchant, 2016; Schwarz et al., 2016). The concept of the placebo effect was well illustrated in the movie *The Wizard of Oz.* The wizard did not *actually* give the tin man a heart, the scarecrow a brain, and the lion courage, but they all felt better nevertheless. In fact, because it is only in the last hundred years or so that medical science has developed a substantial number of treatments that actually have therapeutic efficacy, it is often said that the whole history of medicine before the twentieth century was simply the history of the placebo effect.

We can illustrate the concept of a placebo effect by considering biomedical research, where all studies of new medical procedures must include controls for placebo effects. Typically, if a new drug is being tested on a group of patients, an equivalent group will also be formed and given a pill that does not contain the drug (a placebo). Neither group will know which it is receiving. Thus, when the two groups are compared, the placebo effect—that is, the tendency to feel better when any new treatment is introduced—is controlled for. It would not be sufficient merely to show that a percentage of patients receiving the new drug report relief from their symptoms, because in the absence of a control group it would be impossible to know what percentage is reporting relief due to a placebo effect rather than to the efficacy of the drug itself.

The placebo effect has been found to be 29 percent (of patients reporting satisfactory relief after receiving a placebo) for major depression, 36 percent for duodenal ulcer, 29 percent for migraine headache, and 27 percent for reflux esophagitis (Cho et al., 2005). Placebo effects can be very powerful—so powerful that there have even been reports of people who have become addicted to placebo pills (Roller & Gowan, 2011), needing more and more to maintain their state of health. One study that found that many people who had surgery for torn rotator-cuff tendons reported that their pain was gone even though an MRI indicated that their tendons had not healed (Kolata, 2009). Finally, even so-called open-label placebos have been found to work for lower back pain (Carvalho et al., 2016). An open-label placebo situation is one where people are *told* they are getting a placebo!

It is no doubt that examples like these account for the fact that almost 50 percent of physicians report that they deliberately prescribe placebos (Tilburt et al., 2008). Finally, placebo effects can be modulated by the context of expectation. Research has demonstrated (Waber et al., 2008) that a costly placebo provides more pain relief than a cheap placebo!

Of course, in actual research on drug therapies, the placebo control is not a pill containing nothing but instead is one containing the best currently known agent for the condition. The issue isolated by the experimental comparison is whether the new drug is superior to the best one currently available.

You are given information about placebo effects every time that you take a prescription medication. The next time you receive a prescription medication (or, if you are too healthy, take a look at your grandfather's!) examine carefully the sheet of information that comes with the drug (or look on the drug manufacturer's website) and you will find information about its placebo effects on the medical problem in question. For example, I take a medication called Imitrex (sumatriptan succinate) for relief from migraine headaches. The information sheet accompanying this drug tells me that controlled studies have demonstrated that, at a particular dos-

age level, 57 percent of patients taking this medication receive relief in two hours (I am one of the lucky 57 percent!). But the sheet also tells me that the same studies have shown a placebo effect of 21 percent for this type of headache—21 percent of people receive relief in two hours when their pill is filled with a placebo rather than sumatriptan succinate.

Placebo effects are implicated in all types of psychological therapy (Lilienfeld, 2007). Many people with psychological problems of mild to moderate severity report improvement after receiving psychotherapy. However, controlled studies have demonstrated that some proportion of this recovery rate is due to a combination of placebo effects and the mere passage of time (Driessen et al., 2015). The latter is termed spontaneous remission. Don Redelmeier happens to be a physician who is also a decision scientist (see Lewis, 2017). Early on his training, Redelmeier had been shocked by how often doctors automatically attributed a patient recovery to the specific treatment prescribed without considering other possibilities. In contrast, Redelmeier thought to himself, "So many diseases are self-limiting. They will cure themselves. People who are in distress seek care. When they seek care, physicians feel the need to do something. You put leeches on; the condition improves. And that can propel a lifetime of leeches. You try it and they get better the next day and it is so compelling" (p. 221, Lewis, 2017). Most therapeutic treatments are some unknown combination of an active therapeutic component and a placebo effect. However, it is important to realize that a positive response to a placebo does not mean that a patient's problem was imaginary.

In studies of psychotherapy effectiveness, it is often difficult to determine exactly how to treat the placebo control group, but these complications should not concern us here (see Boot et al., 2013). Instead, it is important to understand why researchers are concerned about separating true therapeutic effects from placebo effects and spontaneous remission. For example, research has shown that psychotherapies do have a positive effect over and above what would be expected purely as the result of a placebo (Engel, 2008; Shadish & Baldwin, 2005). But experiments using placebo controls have demonstrated that merely citing the overall percentage of people who report improvement *vastly* overestimates the degree of improvement that is uniquely due to the particular treatment (Driessen et al., 2015; Tracey et al., 2014). The problem here is that testimonials are just too easy to generate. Even a worthless treatment can appear effective when the base-rate of "feeling better" is so high. In short, placebo effects are potentially occurring whenever a therapeutic intervention is undertaken, *regardless of the efficacy of the intervention.* The problem is that placebo effects are so potent that, no matter how ludicrous the therapy one uses, if it is administered to a large group of people a few will be willing to give a testimonial to its efficacy (the early-morning whack-on-the-head therapy—use it every day and you'll feel better! Send $10.95 for your special, medically tested rubber hammer).

But we really should not joke about such a serious matter. Unwarranted reliance on testimonials and case study evidence may have disastrous consequences. Recall from Chapter 2 that members of a research team that contributed to the modern conceptualization of Tourette syndrome as an organically based disorder pointed out that inappropriate reliance on case study evidence helped to perpetuate the unfalsifiable psychoanalytic explanations of the syndrome that impeded true scientific progress in investigating the nature of the disorder. As retired physician Harriet Hall (2016) says, "personal testimonies are the enemy of proof" (p. 58).

There is a "tough love" lesson in this section, which is why part of the title is: Why Testimonials Are Worthless. People are easily convinced by testimonials. They are easy to process. To think critically about them takes effort, and many people are reluctant to make the cognitive effort. But honest scientists and teachers simply must tell students and the public that this popular mechanism, testimonial evidence, is

indeed worthless. It is not, in fact, evidence of anything. As cognitive psychologist Daniel Levitin (2016) says: "If you bring twenty people with headaches into a laboratory and give them your new miracle headache drug and ten of them get better, you haven't learned anything" (p. 158). You . . . haven't . . . learned . . . *anything*. Tough advice. But true.

The "Vividness" Problem

It is fine to point out how the existence of placebo effects renders testimonials useless as evidence, but we must recognize another obstacle that prevents people from understanding that testimonials cannot be accepted as proof of a claim. Social and cognitive psychologists have studied what is termed the *vividness effect* in human memory and decision making (Slovic, 2007; Slovic & Slovic, 2015; Wang, 2009). When faced with a problem-solving or decision-making situation, people retrieve from memory the information that seems relevant to the situation at hand. Thus, they are more likely to use the facts that are more accessible to solve a problem or make a decision. One factor that strongly affects accessibility is the vividness of information.

The problem is that there is nothing more vivid or compelling than sincere personal testimony that something has occurred or that something is true. The vividness of personal testimony often overshadows other information of much higher reliability. How often have we carefully collected information on different product brands before making a purchase, only to be dissuaded from our choice at the last minute by a chance recommendation of another product by a friend or an advertisement? Car purchases are a typical example. We may have read surveys of thousands of customers in *Consumer Reports* and decided on car X. After consulting the major automotive magazines and confirming that the experts also recommend car X, we feel secure in our decision—until, that is, we meet a friend at a party who knows a friend who knows a friend who bought an X and got a real lemon, spent hundreds on repairs, and would never buy another. Obviously, this single instance should not substantially affect our opinion, which is based on a survey of thousands of owners and the judgment of several experts. Yet how many of us could resist the temptation to overweigh this evidence?

Imagine that you saw the following headline one Friday morning in your newspaper: "Jumbo Jet Crash Kills 393 People." Goodness, you might think, what a horrible accident. What a terrible thing to happen. Imagine, though, that the following Thursday you got up and your newspaper said, "Another Jumbo Jet Disaster: 367 Die." Oh, no, you might think. Not another disaster! How horrible. What in the world is wrong with our air traffic system? And then imagine—please imagine as best you can—getting up the following Friday and seeing in the paper: "Third Tragic Airline Crash: 401 Dead." Not only you but also the nation would be beside itself. A federal investigation would be demanded. Flights would be grounded. Commissions would be appointed. Massive lawsuits would be filed. Magazines would run cover stories. It would be the lead item on television news programs for several days. Television documentaries would explore the issue. The uproar would be tremendous.

But this is not an imaginary problem. It is real. A jumbo jet *does* crash every week. Well, not one jet, but a lot of little jets. Well, not little jets really, but little transportation devices. These devices are called automobiles. And roughly 400 people die in them *each week* in the United States (over 20,000 people each year), enough to fill a jumbo jet (National Highway Traffic Safety Administration, 2014). Worldwide, we lose a 2011 Japanese tsunami's worth of people a *day* to car crashes (p. 346, Lewis, 2017).

So, a jumbo jet's worth of people die in passenger cars on our nation's highways every week, *yet we pay no attention.* Why? Because the "Jumbo Jet's Worth of People Who Die" are not presented to us in a vivid way by the media. Hence, the 400 people who die *each week* in passenger cars (plus the additional 85 who die *each week* on motorcycles) have no vividness for us. We don't talk about them at the dinner table as we do

when a jet goes down and kills a lot of people. We do not debate the safety of car travel as we would the safety of the air traffic system if a jumbo jet crashed every week killing 400 people each time. The 400 are not on the news because they are distributed all over the country and, thus, are a statistical abstraction to most of us. The media do not vividly present to us these 400 deaths because they do not happen in the same place.

Instead, the media present to us (occasionally) a number (e.g., 400 per week). This *should* be enough to get us thinking, but it is not. Driving automobiles is an extremely dangerous activity, compared to almost any other activity in our lives (Galovski et al., 2006; National Safety Council, 2016), yet there has never been a national debate about its risk relative to the benefits involved. Is this an acceptable toll for our lifestyle? We never ask the question because no problem is recognized. No problem is recognized because the cost is not presented to us in a vivid way, as is the cost of airline crashes.

Think of the absurdity of the following example. A friend drives you 20 miles to the airport where you are getting on a plane for a trip of about 750 miles. Your friend is likely to say, "Have a safe trip," as you part. This parting comment turns out to be sadly ironic, because your friend is *three times more likely to die in a car accident on the 20-mile trip back home than you are on your flight of 750 miles.* It is the vividness problem that accounts for the apparent irrationality of person A's wishing person B safety, when it is person A who is in more danger (Sivak & Flannagan, 2003; Smith, 2013).

It was for this reason that the Federal Aviation Administration, although it recommends that infants and toddlers have their own seat on airplanes (with approved child restraint), will not make it a requirement (Associated Press, 2010). The reason they will not is that the FAA worries that, if forced to buy a seat for young children, some parents would choose to drive rather than fly—thus putting the child in much, much more danger than if they were on their parents' laps in an airplane. In our workaday environment, there is no more dangerous place for a child than in a car, yet many parents simply cannot process this fact. In 2016, only 271 people died in commercial plane crashes worldwide, while 1.3 million people died in road crashes (Cheng, 2017).

It is hard to avoid the effects of vividness in our judgments. Take the example of Cornell University, which has the reputation of having a high student suicide rate. We must ask the question of why it has this reputation. We have to ask the question because, statistically, it is not a high suicide school. In fact, its suicide rate is less than half of the national average (Frank, 2007). The reputation has nothing to do with the actual statistics—with the *actual* frequency of suicide at Cornell at all. It has to do with the fact that Cornell is bordered on two sides with deep glaciated gorges—gorges with dramatic bridges spanning them (Frank, 2007). Not surprisingly, the suicides that do occur often take place on these bridges, leading to traffic tie-ups as rescue teams retrieve bodies from the gorge and, most importantly, vivid television footage from the sight of the suicide. A drug overdose leads to no parallel type of media coverage. Cornell's reputation stems from vividness, not statistics.

Misleading personal judgments based on the vividness of media-presented images are widespread in other areas as well. Studies have surveyed parents to see which risks to their children worried them the most (Gardner, 2008; Skenazy, 2009). Parents turned out to be most worried about their children being abducted, an event with a probability of 1 in 600,000. By contrast, the probability of their child being killed in a car crash, which the parents worried about much less, is *dozens of times* more likely (Gardner, 2008). Likewise, children are much more likely to drown in a swimming pool than they are to be abducted and killed by a stranger (Kalb & White, 2010). Of course, the fears of abduction are mostly a media-created worry. Car crashes, accidents, childhood obesity, and suicide at older ages are much more of a threat to our children's well-being than are things like abduction and shark attacks. Because of media-created vividness effects, our risk perception is all out of whack.

The vividness of presentations can even affect the way we interpret scientific evidence itself (Beck, 2010). Several studies have shown that the conclusions of scientific experiments

in cognitive neuroscience were rated as more credible if they contained a brain image summarizing the results instead of a graph depicting the identical outcome (Rhodes et al., 2014; Weisberg et al., 2015). The effect of vivid neuroscience brain images is so extreme and so ubiquitous that journalists have called it "brain porn" (Quart, 2012). Others have called it: "neuromarketeering," "neuromyths," and "biobunk" (Michel, 2015; Voss, 2012).

The Overwhelming Impact of the Single Case

Anecdotal information tends to be weighted too highly by people because of its vividness (Rodriguez et al., 2016). Over-reliance on vivid anecdotal information can negatively impact health behavior and medical decisions. In October of 2011, the U.S. Preventive Services Task Force (USPSTF) recommended that physicians not use the prostate-specific antigen (PSA) test to screen for prostate cancer. The USPSTF's review of the scientific evidence indicated that the harms associated with the test (the side effects associated with unnecessary treatment) outweighed the benefits in mortality (which were tiny at best, and perhaps nonexistent). Psychologists Hal Arkes and Wolfgang Gaissmaier (2012) studied why there was resistance to the recommendation of the USPSTF and concluded that one of the most powerful reasons for resistance was the power of vivid anecdotal evidence.

A well-known example of how people respond differently to vivid anecdotal information comes from the media coverage of the Vietnam War in the mid- to late 1960s. As the war dragged on and the death toll of Americans killed continued without an end in sight, the media took to reporting the weekly number of American service personnel who had been killed that week. Week after week, the figure varied between 200 and 300, and the public, seemingly, became quite accustomed to this report. However, one week a major magazine published a spread, running on for several pages, of the individual pictures of those persons who had died in the previous week. The public was now looking, concretely, at the approximately 250 individual lives that had been lost in a typical week. The result was a major outcry against the toll that the war was taking. The 250 pictures had an effect that the weekly numbers had not had. But we, as a society, must overcome this tendency not to believe numbers—to have to see everything. Most of the complex influences on our society are accurately captured only by numbers. Until the public learns to treat these numerical abstractions of reality as seriously as images, public opinion will be as fickle as the latest image to flicker across the screen.

But it is not only the public that is plagued by the vividness problem. Experienced clinical practitioners in both psychology and medicine struggle all the time with the tendency to have their judgment clouded by the overwhelming impact of the single case. Writer Francine Russo (1999) describes the dilemma of Willie Anderson, an oncologist at the University of Virginia. Anderson is an advocate of controlled experimentation and routinely enrolls his patients in controlled clinical trials, but he still struggles with his own reactions to single, salient cases that have an emotional impact on his decisions. Despite his scientific orientation, he admits that "when it's real people looking you in the eye, you get wrapped up in their hopes and your hopes for their hopes, and it's *hard*" (p. 36). But Anderson knows that sometimes the best thing for his patients is to ignore the "real person looking you in the eye" and go with what the best evidence says. And the best evidence comes from a controlled clinical trial, to be described in Chapter 6.

Media presentations often emphasize vivid unrepresentative examples in their reports. One study of over 100 news stories about student debt found that the student profiled in the story had borrowed an average of $85,000 (Wessel, 2016). But in the year the study was done (2014), only 7 percent of students had borrowed more than $75,000 and two-thirds of all borrowers owed less than $25,000. The students profiled in media stories of student debt are not typical—but they are vivid examples.

Why Vivid Anecdotes and Testimonials Are So Potent

All of these examples of course raise the question of, why vividness? In terms of human cognition, why are vivid testimonials and anecdotes so powerful? Decades of research in cognitive psychology has shown that humans are so-called “cognitive misers” because our natural default is to use the least taxing mental processes (called Type 1 processes) when approaching a problem (Stanovich et al., 2016). This default makes evolutionary sense

because if a problem can be solved by using simple cues, that leaves us extra mental capacity for other tasks. A problem arises, however, when these simple cues are either insufficient or are vastly inferior to the more complex cues available. Such a problem occurs when someone chooses to believe a personal opinion over scientific evidence.

When we evaluate a personal opinion, we automatically engage the evolutionarily old regions of the brain having to do with social engagement. In contrast, understanding scientific evidence, which involves more complex strategic and logical thinking, is a more recent cultural achievement and requires slow and mentally taxing Type 2 processing. From this dual-processing perspective, we can see that there are several ways in which personal opinion might trump scientific thinking. First, because scientific thinking strategies must be acquired, some people may not have learned them. However, even if a person can evaluate scientific evidence, the miserly tendency to default to Type 1 processing may still lead some people to ignore scientific reasoning in the face of a more emotionally persuasive personal opinion. Our tendency to be a cognitive miser often prevents us from carrying to completion the taxing Type 2 processes that are necessary to inhibit Type 1 processing and to substitute statistical thinking. The good news is that scientific and statistical thinking can be practiced to the point of automaticity whereby they themselves become a less taxing mental option. That is what this book is meant to facilitate.

The Amazing Randi: Fighting Fire with Fire

The problems created by reliance on testimonial evidence are ever present. The vividness of such evidence often eclipses more reliable information and obscures understanding. Psychology instructors worry that merely pointing out the logical fallacies of reliance on testimonial evidence is not enough to provide a deep understanding of the pitfalls of these types of data. What else can be done? Is there any other way to get this concept across to people? Fortunately, there is an alternative—an alternative somewhat different from the academic approach. The essence of this approach is to fight vividness with vividness. To hoist testimonials on their own petard! To let testimonials devour themselves with their own absurdity. A practitioner of this approach is the one, the only, the indubitable Amazing Randi!

James Randi is a magician and jack-of-all-trades who has received a MacArthur Foundation “genius” grant. For many years, he has been trying to teach the public some basic skills of critical thinking. The Amazing Randi (his stage name) has done this by exposing the fraud and charlatanism surrounding claims of “psychic” abilities. Randi has used his considerable talents in the service of the public by exposing the fallacies behind ESP, biorhythms, psychic surgery, levitation, and other pseudosciences (Polidoro, 2015; Randi, 2005, 2011; Shermer, 2011).

One of Randi’s minor diversions consists of demonstrating how easy it is to garner testimonial evidence for any preposterous event or vacuous claim. His technique is to let people be swallowed up in a trap set by their own testimonials. On a radio show, Randi demonstrated the basis for the popularity of a pseudoscience called biorhythms (Hines, 2003). One listener agreed to keep a day-by-day diary and compare it with a two-month biorhythm chart that had been prepared especially for her. Two months later, the woman called back to inform the audience that biorhythms should be taken very seriously because her chart was more than 90 percent accurate. Randi had to inform her of the silly mistake made by his secretary, who had sent someone else’s chart to her, rather than her own. However,

the woman did agree to evaluate the correct chart, which would be mailed to her right away, and to call back. A couple of days later, the woman called back, relieved. Her own chart was just as accurate—in fact, even more accurate. On the next show, however, it was discovered that, whoops, another error had been made. The woman had been sent Randi's secretary's chart rather than her own!

Randi's biorhythm scam is actually an example of a phenomenon that has been termed the "P. T. Barnum effect." (Barnum, the famous carnival and circus operator, coined the statement "There's a sucker born every minute.") This effect has been extensively studied by psychologists (Claridge et al., 2008), who have found that the vast majority of individuals will endorse generalized personality summaries as accurate and specific descriptions of themselves. Here is an example taken from Shermer (2005, p. 6):

> You can be a very considerate person, very quick to provide for others, but there are times, if you are honest, when you recognize a selfish streak in yourself. . . . Sometimes you are too honest about your feelings and you reveal too much of yourself. You are good at thinking things through and you like to see proof before you change your mind about anything. When you find yourself in a new situation you are very cautious until you find out what's going on, and then you begin to act with confidence.... You know how to be a good friend. You are able to discipline yourself so that you seem in control to others, but actually you sometimes feel somewhat insecure. You wish you could be a little more popular and at ease in your interpersonal relationships than you are now. You are wise in the ways of the world, a wisdom gained through hard experience rather than book learning.

Large numbers of people find this summary to be a very accurate description of their personality. But very few people spontaneously realize that most other people would also find it indicative of *themselves!* There are well-known sets of statements and phrases (like this example) that most people see as applicable to themselves. Anyone can feed them to a "client" as individualized psychological "analysis" and the client will usually be very impressed by the individualized accuracy of the "personality reading," not knowing that the same reading is being given to everyone. The Barnum effect is, of course, the basis of belief in the accuracy of palm readers and astrologists. The Barnum effect also provides an example of how easy it is to generate testimonials and, of course, shows why they are worthless.

This is exactly what James Randi was trying to do in his little scam described previously—to teach people a lesson about the worthlessness of testimonial evidence. He consistently demonstrates how easy it is to generate testimonials in favor of just about any bogus claim. For this reason, presenting a testimonial in support of a particular claim is meaningless. Only evidence from controlled observations (to be described in Chapter 6) is sufficient to actually *test* a claim.

Testimonials Open the Door to Pseudoscience

It is sometimes claimed that pseudosciences like parapsychology, astrology, biorhythms, and fortune-telling are simply a way to have a little fun, that they really do no harm. After all, why should we care? Isn't it just a case of a few people engaging in wishful thinking and a few others making a couple of bucks out of them? In fact, a complete examination of the problem reveals that the harm done to society by the prevalence of pseudosciences is more widespread than is generally believed.

First, people tend not to think about what economists call *opportunity costs.* When you take time to do one thing, you have lost the time to do something else. You have

lost the opportunity to spend your time otherwise. When you spend money on one thing you no longer have the money to do something else—you have lost the opportunity to spend it otherwise. Pseudosciences have massive opportunity costs. When people spend time (and money) on pseudosciences they gain nothing and they waste time that might have been spent on more productive endeavors.

A complete examination of the problem reveals that the harm done to society by the prevalence of pseudosciences is more widespread than is generally believed. And the costs extend beyond opportunity costs. In a complex, technological society, the influence of pseudoscience can be propagated by decisions that affect thousands of other people. That is, you may be affected by pseudoscientific beliefs even if you do not share those beliefs. For example, one-third of Americans are drinking unfluoridated water despite voluminous scientific evidence that fluoridation can significantly reduce tooth decay (Beck, 2008; Griffin et al., 2007; Singh et al., 2007). The Centers for Disease Control estimate that for every dollar spent on fluoridation, $38 in dental treatment costs are saved (Brody, 2012). Nevertheless, millions of Americans in areas without fluoridation are suffering needless cavities because their neighbors are in the grip of pseudoscientific conspiracy theories about the harmful effects of fluoridation. Small groups of people with these pseudoscientific beliefs have pressured various communities to keep fluoridation out and have thus denied its benefits to everyone who lives near them. In short, the pseudoscientific beliefs of the few have negatively affected the many.

We are all affected in numerous ways when pseudoscientific beliefs permeate society—even if we do not subscribe to the beliefs (Shermer, 2005; Stanovich, 2009). For example, police departments hire psychics to help with investigations even though research has shown that this practice has no effectiveness (Radford, 2010; Shaffer & Jadwiszczok, 2010). There is not a single documented case of psychic information being used to successfully find a missing person (Radford, 2009).

Pseudosciences such as astrology are now large industries, involving newspaper columns, radio shows, book publishing, magazine articles, websites, social media, and other means of dissemination. They are lucrative businesses, and the incomes of thousands of individuals depend on their public acceptance. These businesses cost the public millions of dollars.

And there are other costs in addition to opportunity costs and monetary costs. Thirty-three percent of dietary supplements and herbal products contain ingredients that are not listed on their product labels, some potentially dangerous (Newmaster et al., 2013). Millions of dollars have been spent by our government to promote complementary and alternative medicine, even though the techniques being promoted have not been proven effective in scientifically controlled experiments (Dorlo et al., 2015; Mielczarek & Engler, 2013; Swan et al., 2015). That procedures of alternative medicine have not been scientifically verified should not surprise us, says cancer survivor Gideon Burrows in his book titled *This Book Won't Cure Your Cancer* (2015). Burrows, who survived because of conventional cancer treatment (Hall, 2016), reminds us that *of course* alternative medicine has not been verified by scientific studies. If it had been, we would simply call it *medicine*!

Some associations and organizations have been more aggressive than psychology in rooting out pseudoscience. In 2007 the Federal Trade Commission (FTC) levied multimillion dollar fines against four diet-drug marketers who sold by using infomercials and celebrity endorsements. In announcing the fines, the FTC Chairwoman Deborah Platt Majoras tried to educate the public by stating that "Testimonials from individuals are not a substitute for science, and that's what Americans need to understand" (de la Cruz, 2007, p. A10). Likewise, medical associations have been more aggressive than psychology in attacking pseudoscience and dissociating legitimate medical practice from the illegitimate. Consider the guidelines published by the Arthritis Foundation and cited by the House Committee on Aging for spotting the

unscrupulous promoter (US Congress, 1984):

1. He may offer a "special" or "secret" formula or device for "curing" arthritis.
2. He advertises. He uses "case histories" and testimonials from satisfied "patients."

3. He may promise (or imply) a quick or easy cure.
4. He may claim to know the cause of arthritis and talk about "cleansing" your body of "poisons" and "pepping up" your health. He may say surgery, X-rays, and drugs prescribed by a physician are unnecessary.
5. He may accuse the "medical establishment" of deliberately thwarting progress, or of persecuting him . . . but he doesn't allow his method to be tested in tried and proven ways.

This list could also serve as a guide for spotting fraudulent psychological treatments and claims. Note, of course, point 2, which is the focus of this chapter. But also note that points 1 and 5 illustrate the importance of something discussed earlier: Science is public. In addition to putting forth testimonials as "proof," the practitioners of pseudoscience often try to circumvent the public verifiability criterion of science by charging that there is a conspiracy to suppress their "knowledge." They use this as justification for going straight to the media with their "findings" rather than submitting their work to the normal scientific publication processes.

A caution that might be added to the list above is to watch out for situations where someone seems to be offering an outcome that allows one to escape well-established trade-offs. For example, in investing it is well known that risk is related to reward (larger investment returns require taking more risk). In dieting, it is well known that long-term weight reduction depends on long-term changes in caloric intake. Regarding educational interventions, it is well known that the more long-lasting educational gains come from intensive intervention programs of long duration. In short: financial returns and risk trade-off; weight loss and caloric intake trade-off; and learning gains and intensity of intervention trade-off. People pushing pseudoscientific ideas in these areas will invariably claim that they can break free of these trade-offs—that you can have high monetary returns without risk; that you can lose weight and still eat all you want; and that educational achievement can be significantly altered by short-term interventions. You can be confident that claims that such fundamental trade-offs have been avoided are bogus. For example, you can rest assured that a product called Baby Einstein can't deliver what its name implies (Bronson & Merryman, 2009; DeLoache et al., 2010).

It is important to realize that television, websites, and the print media will publicize virtually any outlandish claim in the area of psychology if they think there is an audience for it, no matter how much the claim is contradicted by the available evidence. The media present a mish-mosh of claims and experts—mixing together legitimate scientists with pseudoscientific charlatans. At the outset of this chapter, I mentioned the self-help television shows of the 1990s and 2000s and shows like the *Oprah Winfrey Show*—popular for two decades on television. To be fair, it must be said that that show often presented credentialed professionals who gave the audience good information on a variety of topics from breast cancer to personal finance. But mixed into the show, with often no way to tell the difference, were the most outrageous charlatans (Gardner, 2010)—often speaking on some of the same topics. For example, Oprah publicized the alternative therapies of someone who used Tarot cards to diagnose illness and who viewed women's thyroid problems as the result of "energy blockage in the throat region" that resulted from "a lifetime of swallowing words one is aching to say" (Kosova & Wingert, 2009, p. 59).

People can fail to take advantage of the real remedies available to them because they become involved in pseudosciences. Many sick individuals delay

getting medically appropriate treatment because they waste time chasing bogus cures. Renowned computer entrepreneur Steve Jobs ignored his doctors after being told of his pancreatic cancer and delayed surgery for nine months while he pursued unproven fruit diets, consulted a psychic, and received bogus hydrotherapy (Isaacson, 2011; Shermer, 2012).

A clear example of how we are all hurt when pseudoscientific beliefs spread is provided by the theory (first put forth in the early 1990s and continuing to this day) that autism is connected to the early vaccination of children. This theory is false. It is contradicted by voluminous evidence (Grant, 2011; Honda et al., 2005; Nyhan et al., 2014; Offit, 2011; Taylor, 2006), but no reader of this chapter should be surprised at how the belief arose. Many children are diagnosed with autism around the time of their first vaccinations and many begin to show clearly discernible signs of the condition (delayed language acquisition, difficulties in reciprocal social interaction, and a restricted repertoire of activities) around this time. Not surprisingly given that there are thousands of children with this condition, some parents become fully aware of their child's difficulties (either through diagnosis or increased awareness based on their own observations) shortly after the child receives a vaccination. These parents then provide the vivid and heartfelt testimonials that there must be a connection between their child's condition and the vaccination. However, many different experimental and epidemiological studies have converged (see Chapter 8) on the conclusion that no such connection exists (Deer, 2011; Offit, 2011; Randi, 2017).

This pseudoscientific belief, though, had more costs than just the opportunity costs to the parents and children involved. The false belief in a connection spawned an antivaccination movement. As a result, immunization rates have decreased, many more children have been hospitalized with measles than would have been the case otherwise, and some have died (Brody, 2015; Cheng, 2013; Nisbet, 2016). Again, the lesson is that in an interconnected society, your neighbor's pseudoscientific belief might affect you even if you reject the belief yourself.

Political leaders, when they believe in pseudoscience, can inflict the consequences of their beliefs on thousands of people. The second postapartheid President of South Africa, Thabo Mbeki, rejected the scientific consensus that AIDS was caused by a virus (Pigliucci, 2010). The neighboring countries of Botswana and Namibia gave an antiretroviral to their citizens infected with HIV, but South Africa did not. Instead, Mbeki set up a commission of AIDS denialists. They recommended that HIV testing be stopped and that AIDS patients be treated with massage therapy, music therapy, yoga, and a naturopathic diet of garlic and beet root (Mielczarek & Engler, 2014; Nattrass, 2012). It is estimated that the rejection of the anti-retrovirals caused 330,000 South Africans to endure premature deaths.

When I am speaking on the topic of how vivid testimonials mislead, at some point in my lecture someone always asks a very relevant question: "Haven't you just employed vivid cases to illustrate a point—what you said shouldn't be done?" This is a good question and it allows me to elaborate on some of the subtleties involved in the argument of this chapter. The answer to the question is that yes, I *have* employed vivid cases to illustrate a point. To *illustrate* the point—but not to *prove* it. The key issue here is to distinguish two things: (1) the claim being made and (2) the communication of the claim. For each one we could ask, is its basis a vivid testimonial, yes or no? This yields four possible situations:

a. a claim based on vivid testimonials communicated by vivid testimonials
b. a claim based on vivid testimonials communicated without testimonials
c. a claim based on evidence other than testimonials communicated by vivid testimonials
d. a claim based on evidence other than testimonials communicated without testimonials

testimonials

Some of the discussion in this chapter falls into category c: claims based on evidence other than testimonials communicated by vivid testimonials. For example, I cite much nontestimonial evidence throughout the chapter to establish claims such as the following: Case study evidence cannot be used to establish a causal influence, vivid examples are overweighted in people's judgments, pseudoscience is costly, and so on.

I present in the citations and the reference list the public evidence for each of these claims. Nonetheless, for communicative purposes I have used some vivid cases to draw attention to these claims and to make them memorable. The key point, though, is that the claims themselves are supported by more than vivid testimonial evidence. So, for example, I have used some vivid examples to demonstrate the fact that vivid examples are overweighted in people's judgments. But the real evidence for the claim that vivid examples are overweighted in people's judgments is in the peer-reviewed scientific evidence I have cited (e.g., Obrecht et al., 2009; Slovic & Slovic, 2015; Wang, 2009).

So, to return to the main point of this section, and to summarize it: The spread of pseudoscience is quite costly. And nothing fosters the spread of pseudosciences more than confusion about what type of evidence does and does not justify belief in a claim about a phenomenon. By providing readily available support for virtually any claim and by the impact that they have when used, testimonials open the door to the development of and the belief in pseudosciences. There is no more important rule for the consumer of psychological information than to beware of them. In the next several chapters, we shall see what type of evidence *is* required to justify knowledge claims.

It is important to learn the scientific reasoning principles in the remaining chapters of this book, because there is no shortcut to assessing scientific credibility. For example, it is sadly true that universities themselves are no guarantee of quality control. Many university courses are full of pseudoscientific information (Lilienfeld, 2017; Novella, 2012), and many courses (indeed, sometimes entire departments!) in universities are explicitly anti-science in their approach. This sometimes includes courses in psychology departments. Not all courses in psychology departments adhere to the principles articulated in this book. Some university departments are more engaged in political advocacy than they are true inquiry (Furedi, 2017; Lilienfeld, 2017; Lukianoff, 2012; Lukianoff & Haidt, 2015; Otto, 2016). Students need to learn to think scientifically themselves because, outside of the STEM disciplines, there is no guarantee that your university instructors are actually demonstrating scientific thinking styles.

Summary

Case study and testimonial evidence is useful in psychology (and other sciences) in the very earliest stages of an investigation, when it is important to find interesting phenomena and important variables to examine further. As useful as case study evidence is in the early pretheoretical stages of scientific investigation, it is virtually useless in the later stages, when theories are being put to specific tests. This is because, as an isolated phenomenon, the outcome of a case study leaves many alternative explanations. One way to understand why case studies and testimonial evidence are useless for theory testing is to consider the placebo effect. The placebo effect is the tendency of people to report that any treatment has helped them, regardless of whether the treatment had a real therapeutic element. The existence of placebo effects makes it impossible to prove the effectiveness of a psychological (or medical) treatment by producing testimonials to its effectiveness. The reason is that the placebo effect guarantees that no matter what the treatment, it will be possible to produce testimonial evidence to its effectiveness.

Despite the uselessness of testimonial evidence in theory testing, psychological research has indicated that such evidence is often weighted quite heavily by people because of the vividness effect: People overweigh evidence that is more vivid and, hence, more retrievable from memory. One thing that is particularly vivid for most people is testimonial evidence. The result is an overreliance on such evidence in the justification of specific psychological claims. In fact, testimonial and case study evidence cannot be used to justify general theoretical claims.

Chapter 5

Correlation and Causation: Birth Control by the Toaster Method

Learning Objectives

5.1 Explain the third variable problem in correlational research
5.2 Explain the directionality problem in correlational research
5.3 Outline how selection bias leads to spurious correlations

Many years ago, a large-scale study of the factors related to the use of contraceptive devices was conducted in Taiwan. A research team of social scientists collected data on a wide range of behavioral and environmental variables. The researchers were interested in seeing what variables best predicted the adoption of birth control methods. After collecting the data, they found that the one variable most strongly related to contraceptive use was the number of electrical appliances (toasters, fans, etc.) in the home (Li, 1975).

This result probably does not tempt you to propose that the teenage pregnancy problem should be dealt with by passing out free toasters in high schools. But why *aren't* you tempted to think so? The correlation between appliances and contraceptive use was indeed strong, and this variable was the single best predictor among the many variables that were measured. Your reply, I hope, will be that it is not the strength but the *nature* of the relationship that is relevant. Starting a free toaster program would imply the belief that toasters *cause* people to use contraceptives. The fact that we view this suggestion as absurd means that, at least in clear-cut cases such as this, we recognize that two variables may be associated without having a causal relationship.

In this example, we can guess that the relationship exists because contraceptive use and the number of electrical appliances in the home are linked through some other variable that relates to both. Socioeconomic status (SES) would be one likely candidate for a mediating variable. We know that SES is related to contraceptive use. All we need now is the fact that families at higher socioeconomic levels tend to

have more electrical appliances in their homes, and we have the linkage. Of course, other variables may mediate this correlation. However, the point is that, no matter how strong the correlation is between the number of toasters and contraceptive use, the relationship does not indicate a causal connection.

The contraceptive example makes it very easy to understand the fundamental principle of this chapter: The presence of a correlation does not necessarily imply causation. In this chapter, we will discuss two problems that prevent the drawing of a causal inference: the third-variable problem and the directionality problem. The toaster-contraceptive study is an example of the third-variable problem.

The Third-Variable Problem

Sometimes the third variable causing a misleading linkage between two other variables is pretty easy to see. If I told you that, across the 365 days of the year, at all of the beach resorts in America, there is a correlation between the number of ice cream cones sold and the number of drownings. The more ice cream sold, the more drownings there were. Here, it is easy to see that the association comes about not because people fill their stomachs with ice cream and that that makes them drown when they go in the water. Instead, a third variable, the temperature, links these two variables. On hot days, when there are a lot of people eating ice cream, there are also a lot of people swimming, and the more people there are out swimming, the more there are who will drown.

The limitations of correlational evidence are not always so easy to recognize as the ice cream and toaster examples. When the causal link seems obvious to us, when we have a strong preexisting bias, or when our interpretations become dominated by our theoretical orientation, it is tempting to treat correlations as evidence of causation.

In the early 1900s, thousands of Americans in the South suffered and died of a disease called *pellagra*. Characterized by dizziness, lethargy, running sores, vomiting, and severe diarrhea, the disease was thought to be infectious and to be caused by a living microorganism of "unknown origin." It is not surprising, then, that many physicians of the National Association for the Study of Pellagra were impressed by evidence that the disease was linked to sanitary conditions. It seemed that homes in Spartanburg, South Carolina, that were free of pellagra invariably had inside plumbing and good sewerage. By contrast, the homes of pellagra victims often had inferior sewerage. This correlation coincided quite well with the idea of an infectious disease transmitted, because of poor sanitary conditions, via the excrement of pellagra victims.

One physician who doubted this interpretation was Joseph Goldberger, who, at the direction of the surgeon general of the United States, had conducted several investigations of pellagra. Goldberger thought that pellagra was caused by inadequate diet. Many victims had lived on high-carbohydrate, extremely low-protein diets, characterized by small amounts of meat, eggs, and milk and large amounts of corn, grits, and mush. Goldberger thought that the correlation between sewage conditions and pellagra did not reflect a causal relationship in either direction (much as in the toaster–birth control example). Goldberger thought that the correlation arose because families with sanitary plumbing were likely to be economically advantaged. This economic discrepancy would also be reflected in their diets, which would contain more animal protein.

But wait a minute! Why should Goldberger get away with his causal inference? After all, both sides were just sitting there with their correlations, Goldberger

with pellagra and diet and the other physicians with pellagra and sanitation. Why shouldn't the association's physicians be able to say that Goldberger's correlation was equally misleading? Why was he justified in rejecting the hypothesis that an infectious organism was transmitted through the excrement of pellagra victims because of inadequate sewage disposal? Well, the reason Goldberger was justified has to do with one small detail that I neglected to mention: Goldberger had eaten the excrement of pellagra victims.

Why Goldberger's Evidence Was Better

Goldberger had a type of evidence (a controlled manipulation, discussed further in the next chapter) that is derived when the investigator, instead of simply observing correlations, actually manipulates the critical variable. This approach often involves setting up special conditions that rarely occur naturally—and to call Goldberger's special conditions unnatural is an understatement!

Confident that pellagra was not contagious and not transmitted by the bodily fluids of the victims, Goldberger had himself injected with the blood of a victim. He inserted throat and nose secretions from a victim into his own mouth. According to Bronfenbrenner and Mahoney (1975), two researchers describing Goldberger's efforts, he and his assistants even ate dough balls that contained the urine and feces of pellagra victims! Despite all of these extreme interventions, neither Goldberger nor the other volunteers came down with pellagra. In short, Goldberger had created the conditions necessary for the infectious transmission of the disease, and nothing had happened.

Goldberger had now manipulated the causal mechanism suggested by others and had shown that it was ineffective, but it was still necessary to test his own causal mechanism. Goldberger got two groups of prisoners from a state prison farm who were free of pellagra to volunteer for his experiment. One group was given the high-carbohydrate, low-protein diet that he suspected was the cause of pellagra, while the other group received a more balanced diet. Within five months, the low-protein group was ravaged by pellagra, while the other group showed no signs of the disease. After a long struggle, Goldberger's hypothesis was eventually accepted because it matched the empirical evidence better than any other.

The history of pellagra illustrates the human cost of basing social and economic policy on mistaken inferences from correlational studies. This is not to say that we should never use correlational evidence. Quite the contrary. In many instances, it is all we have to work with (see Chapter 8), and in some cases, it is all we need (for instance, when prediction, rather than determination of cause, is the goal). Scientists often have to use incomplete knowledge to solve problems. The important thing is that we approach correlational evidence with a certain skepticism. Examples such as the pellagra–sewage case occur with considerable frequency in all areas of psychology. The example illustrates what is termed the *third-variable problem*: the fact that the correlation between the two variables—in this case, pellagra incidence and sewage conditions—may not indicate a direct causal path between them but may arise because both variables are related to a third variable that has not even been measured. Pellagra incidence is related to SES (and to diet—the real causal variable) and SES is also related to sewerage quality. Correlations like that between sewage and pellagra are often called spurious correlations: correlations that arise not because a causal link exists between the two variables that are measured, but because both variables are related to a third variable (or just show a chance relationship, see Vigen, 2015).

It is sometimes easy to fall into the trap of ignoring possible third variables. When we see a study showing correlations between parenting behaviors and their children's psychological characteristics, it is tempting to automatically think that the parenting behaviors *determined* (that is, caused) the children's psychological characteristics. But this automatic tendency is wrong because it ignores the genetic connection between the parents and their children—a third variable that may be responsible for the parent-child correlations (McAdams et al., 2014;

Plomin et al., 2016).

Let's consider a contemporary example of a third variable problem. For decades, debates have raged over the relative efficacy of public and private schools. Some of the conclusions drawn in this debate vividly demonstrate the perils of inferring causation from correlational evidence. The question of the efficacy of private versus public schools is an empirical problem that can be attacked with the investigative

methods of the social sciences. This is not to imply that it is an easy problem, only that it is a scientific problem, and potentially solvable. All advocates of the superiority of private schools implicitly recognize this, because at the crux of their arguments is an empirical fact: Student achievement in private schools exceeds that in public schools. This fact is not in dispute—educational statistics are plentiful and largely consistent across various studies. The problem is the use of these achievement data to conclude that the education received in private schools *causes* the superior test scores.

The outcome of educational testing is a function of many different variables, all of which are correlated. In order to evaluate the relative efficacy of public schools and private schools, we need more complex statistics than merely the relationship between the type of school attended and school achievement. For example, educational achievement is related to many different indicators of family background, such as parental education, number of parents in the home, SES, the number of books in the home, and other factors. These characteristics are also related to the probability of sending a child to a private school. Thus, family background is a potential third variable that may affect the relationship between academic achievement and the type of school. In short, the relationship may have nothing to do with the effectiveness of private schools but may be the result of the fact that economically advantaged children do better academically and are more likely to attend private schools.

Fortunately, there exist complex correlational statistics such as *multiple regression, partial correlation*, and *path analysis* that were designed to deal with problems such as this one (Wheelan, 2013). These statistics allow the correlation between two variables to be recalculated after the influence of other variables is removed, or "factored out" or "partialed out." Using these more complex correlational techniques, researchers have analyzed a large set of educational statistics on high school students (Carnoy et al., 2005). They found that, after variables reflecting the students' home backgrounds and general mental ability were factored out, there was little relationship between school achievement and the type of school attended. Academic achievement is linked to private school attendance not primarily because of any direct causal mechanism, but because the family background and the general cognitive level of students in private schools are different from those of children in public schools.

It requires similar statistical techniques (regression, partial correlation) to untangle the relationship between happiness and longevity. There is indeed a positive correlation: happier people live longer. But from this correlation alone, we cannot conclude that happiness is the cause of greater longevity. In fact, some studies have shown (by using the statistical regression technique) that the correlation between happiness and longevity disappears when levels of health are controlled (Liu et al., 2016; Vyse, 2016a).

The complex correlational statistics that allow us to partial out the effects of a third variable do not always reduce the magnitude of the original correlation. Sometimes the original correlation between two variables remains even after the partialing out of the third variable, and this result itself can be informative. Such an outcome indicates that the original correlation was not due to a spurious relationship with that particular third variable. Of course, it does not remove the possibility of a spurious relationship due to some other variable. For example, it turns out that the violent crime rate in the United States is higher in the southern states than in the northern states. Anderson and Anderson (1996) tested what has been called the *heat hypothesis*—that "uncomfortably warm temperatures produce increases in aggressive

motives and (sometimes) aggressive behavior" (p. 740). Not surprisingly, they did find a correlation between the average temperature in a city and its violent crime rate. What gives the heat hypothesis more credence, however, is that they found that the correlation between temperature and violent crime remained significant even after variables such as unemployment rate, per capita income, poverty rate, education, population size, median age of population, and several other variables were statistically controlled (see also, Larrick et al., 2011; Plante & Anderson, 2017).

The Directionality Problem

There is no excuse for making causal inferences on the basis of correlational evidence when it is possible to manipulate variables in a way that would legitimately justify a causal inference. Yet this is a distressingly common occurrence when psychological issues are involved. A well-known example in the area of educational psychology illustrates this point quite well.

Since the beginning of the scientific study of reading about a hundred years ago, researchers have known that there is a correlation between eye movement patterns and reading ability. Poorer readers make more erratic movements, display more regressions (movements from right to left), and make more fixations (stops) per line of text. On the basis of this correlation, some educators hypothesized that deficient oculomotor skills were the cause of reading problems, and many eye movement–training programs were developed and administered to elementary school children. These programs were instituted long before it was ascertained whether the correlation really indicated that erratic eye movements *caused* poor reading.

It is now known that the eye movement–reading-ability correlation reflects a causal relationship that runs in exactly the opposite direction (Rayner et al., 2012). Erratic eye movements do not cause reading problems. Instead, slow recognition of words and difficulties with comprehension lead to erratic eye movements. When children are taught to recognize words efficiently and to comprehend better, their eye movements change. Training children's eye movements does nothing to improve their reading comprehension.

For more than a decade now, research has clearly pointed to word decoding and a language problem in phonological processing as the sources of reading problems (Cunningham & Zibulsky, 2014; Hulme & Snowling, 2013; Seidenberg, 2017; Willingham, 2017). Very few cases of reading disability are due to difficulties in the area of eye movement patterns. Many school districts still have in their storage basements the dusty "eye movement trainers" that represent thousands of dollars of equipment money wasted because of the temptation to see a correlation as proof of a causal hypothesis.

Consider another somewhat similar example. An extremely popular hypothesis in the fields of education and counseling psychology has been that school achievement problems, drug abuse problems, teenage pregnancy, bullying, and many other problem behaviors are the result of low self-esteem. It was assumed that the causal direction of the linkage was obvious: Low self-esteem led to problem behaviors, and high self-esteem led to high educational achievement and accomplishments in other domains. This assumption of causal direction provided the motivation for many educational programs for improving self-esteem. The problem here was the same as that in the eye movement example: An assumption of causal direction was made from the mere existence of a correlation. It turns out that the relationship between self-esteem and school achievement, if it exists at all, is more likely to be in the opposite direction: Superior accomplishment in school (and in other aspects of life) leads to high self-esteem, rather than the reverse (Krueger et al., 2008; Lilienfeld et al., 2012).

An example often pointed to in research methodology textbooks concerns a group of islanders in the New Hebrides who believed that lice made people healthy, because healthy islanders often had a lot of lice and ill islanders often did not. It turns out that almost all islanders had *some* lice most of the time. As the lice intensify, they cause a fever which kills them. Unhealthy people more quickly got the fever and became free

of lice (Mazur, 2016). This created the observation that healthy people had more lice than unhealthy people. But the causality ran in the other direction: poor health led to less lice (and better health led to more lice) rather than lice leading to health.

Problems of determining the direction of causation are common in psychological research. For example, psychologist Jonathan Haidt (2006) has discussed research showing that there is a correlation between altruism and happiness. There is research showing, for instance, that people who do volunteer work are happier than those who

do not. Of course, it was necessary to make sure that a third variable wasn't accounting for the link between altruism and happiness. Once third variables were eliminated, it was necessary to determine the *direction* of the linkage: Was it that happiness caused people to be altruistic or was it that acts of altruism made people happy ("it is more blessed to give than to receive")? When the proper controlled studies were done, using the logic of the true experiment to be described in Chapter 6, it was found that there was a causal relationship running in *both* directions: Being happy makes people more altruistic, and performing altruistic acts makes people happier.

Psychologists Chris Chabris and Dan Simons (2013) discussed a study in which researchers surveyed 2881 people in 228 census tracts and found that the census tracts with more outdoor food advertising had people who were more obese. Chabris and Simons discuss how the study was presented as if it were correct to assume that the food advertising was affecting people and making them more obese. Readers of this book will by now have thought of the alternative interpretation that runs in the opposite direction: advertisers might place more ads in neighborhoods that are high consumers of their food items.

Earlier we warned about the tendency to, when seeing a study showing correlations between parenting behaviors and children's psychological characteristics, automatically think that the parenting behaviors caused the children's psychological characteristics. We pointed out that the genetic connection between the parents and their children might be a third variable responsible for the parent-child correlations. But, additionally, there may be a directionality problem as well: the child's behaviors might be *evoking* parental responses (Jaffee et al., 2012). The direction of causation might actually be from child to parent.

Our discussion thus far has identified the two major classes of ambiguity present in a simple correlation between two variables. One is called the *directionality problem* and is illustrated by the eye movement and self-esteem examples. Before immediately concluding that a correlation between variable A and variable B is due to changes in A causing changes in B, we must first recognize that the direction of causation may be the opposite, that is, from B to A. The second problem is the third-variable problem, and it is illustrated by the pellagra example (and the toaster–birth control and private-school–achievement examples). The correlation between the two variables may not indicate a causal path in either direction but may arise because both variables are related to a third variable.

Selection Bias

The term *self selection bias* refers to situations where people select themselves into a particular group rather than being randomly assigned (see Chapter 6) to the group. Self selection creates spurious correlations between personal variables and environmental characteristics—correlations that do not indicate a causal relationship. The correlations arise because people with certain behavioral/biological characteristics have chosen certain types of environments, not because the environments *caused* the behavioral/biological characteristics. Self selection is better explained by considering some examples.

Let's look at a straightforward example that illustrates the importance of selection factors in creating spurious correlations: Quickly, name a state with an above-

average incidence of deaths due to respiratory illness. One answer to this question would be, of course, Arizona....What?.... Wait a minute! Arizona has clean air, doesn't it? Does the smog of Los Angeles spread that far? Has the suburban sprawl of Phoenix become that bad? No, it can't be. Let's slow down a minute. Maybe Arizona *does* have good air. And maybe people with respiratory illnesses tend to move there. And then they die. There you have it. A situation has arisen in which, if we're not careful, we may be led to think that Arizona's air is killing people.

However, selection factors are not always so easy to discern. They are often overlooked, particularly when there is a preexisting desire to see a certain type of causal link. Tempting correlational evidence combined with a preexisting bias may deceive even the best of minds. Let's consider some specific cases.

An example from clinical psychology demonstrates how tricky and "perverse" the selection bias problem can be. It has sometimes been demonstrated that the cure rate for various addictive-appetite problems such as obesity, heroin use, and cigarette smoking is lower for those who have had psychotherapy than for those who have not. The reason, you will be glad to know, is not that psychotherapy makes addictive behavior more resistant to change. It is that, among those who seek psychotherapy, the disorder is more intractable (Satel & Lilienfeld, 2013), and self-cures have been ineffective. In short, "hard cases" seek psychotherapy more than "easy cases." This "hard case" self selection bias is so ubiquitous that it has been called the clinician's illusion (Satel & Lilienfeld, 2013)—the illusion being that clinicians tend to overgeneralize the characteristics of the extreme cases they see to the much larger population of milder cases that are less likely to come into contact with a clinician.

This type of self selection bias comes into play when organizations or governments launch so-called scorecards to rate physicians. New York did this some years ago when they started publishing the mortality rates of cardiologists (Wheelan, 2013). The problem was that a cardiologist could simply boost their ratings by seeking out easy cases and avoiding the most difficult cases!

Jumping to conclusions when selection effects are present can lead us to make bad real-world choices. Many women were once encouraged to take hormone replacement therapy (HRT) after menopause because of reports that it lowered the probability of heart disease. But the early studies that had indicated this had simply compared groups of women who had chosen to take HRT (i.e., who self-selected the treatment) with those who had not chosen to take HRT. However, true experiments (using random assignment, see Chapter 6) conducted later on found that HRT actually did not reduce the likelihood of heart disease at all (Bluming & Tavris, 2009; Seethaler, 2009). The earlier studies involving self-selected samples had *seemed* to indicate that it did because women who chose to have HRT were more physically active, less obese, and less likely to smoke than women who did not choose HRT.

Selection bias can lead to some surprising conclusions. During World War II an analyst was trying to determine where to place extra armor on aircraft based on the pattern of bullet holes in the returning planes (Ellenberg, 2014). His decision was to put the extra armor in the places that were free of bullet holes on the returning aircraft that he analyzed. He did *not* put the extra armor in the places where there were a lot of bullet holes. His reasoning was that the planes had probably been pretty uniformly hit with bullets. Where he found the bullet holes on the returning aircraft told him that, in those places, the plane could be hit and still return. Those areas that were free of bullet holes on returning planes had probably been hit—but planes hit *there* did not return. Hence, it was the places on the returning planes *without* bullet holes that needed more armor!

It is easy to use selection effects to "set up" people to make a causal inference. How about this one: Republicans enjoy sex more than Democrats. It's an absolute fact. Statistics show that the average Republican voter is more satisfied with their sex lives than the average Democratic voter (Blastland & Dilnot, 2009). What is it about

Republicanism that makes people sexier?

OK, you guessed it. That's not right. Politics doesn't change anyone's sex life. What accounts for the data, then? Two things. First, married people vote Republican more than single people. Second, surveys show that married people report more satisfaction with their sex lives than single people. Republicanism doesn't change anyone's sex life; it's just that a demographic group (married people) who have higher satisfaction levels are more prone to vote Republican.

Examples such as the "sexy Republican" show us how careful we have to be when selection effects might be operating. Economist Steven Landsburg (2007) demonstrates how much of the data showing productivity tied to the use of technology might be interpreted as causal when in fact it is only correlational data containing selection effects. Within corporations, it is often the most productive employees who are given the most advanced technology. Thus, when a correlation is calculated, productivity will be correlated with technology use. But it is not that the technology improved the performance of these employees, because they were *already* more productive before they received the advanced technology.

An important real-life health issue that implicates selection effects strongly is the debate about the health outcomes of alcohol consumption. Numerous studies have found that moderate drinkers have better health outcomes than not only frequent drinkers but also abstainers as well (Rabin, 2009). Aware of selection effects, neither you nor I will be tempted to tell anyone abstaining from alcohol that they would improve their health by drinking a little. This is because people self select themselves into drinking groups by deciding how much to drink. As Rabin (2009) explains, it has been found that moderate drinkers are moderate in everything they do. They exercise moderately and eat moderately. They tend to do a lot of things right. So of course the problem is that we do not know whether it is the moderate drinking itself that leads to positive health outcomes or whether it is all of the other good characteristics of the moderate drinking group (their exercise levels, diet, etc.). Because of selection effects, we cannot say that the moderate drinking itself is the cause.

It is likewise with some correlational studies that have shown that wine drinkers have better health outcomes than do beer drinkers or liquor/cocktail drinkers (University of California, 2015a). The problem is that wine drinkers generally have healthier habits than beer or liquor drinkers and are different demographically. Wine drinkers smoke less, for instance, and they are more educated and affluent. When studies have employed statistical regression techniques to control for these factors, the association of positive health outcomes with wine drinking disappears.

In short, the consumer's rule for this chapter is simple: Be on the lookout for instances of selection bias, and avoid inferring causation when data are only correlational. It is true that complex correlational designs do exist that allow limited causal inferences. It is also true that correlational evidence is helpful in demonstrating convergence on a hypothesis (see Chapter 8). Nevertheless, it is probably better for the consumer to err on the side of skepticism than to be deceived by correlational relationships that falsely imply causation.

Summary

The central point of this chapter was to convey that the mere existence of a relationship between two variables does not guarantee that changes in one are causing changes in the other. Correlation does not imply causation. Two problems in interpreting correlational relationships were discussed. In the third-variable problem, the correlation between the two

The other thing that makes the interpretation of correlations difficult is the existence of the directionality problem: the fact that even if two variables are causally related, the direction of that relationship is not indicated by the mere presence of the correlation.

Selection bias is the reason for many spurious rela-

variables may not indicate a direct causal path between them but instead may arise because both variables are related to a third variable that has not even been measured. If, in fact, the potential third variable has been measured, correlational statistics such as partial correlation (to be discussed again in Chapter 8) can be used to assess whether that third variable is determining the relationship.

tionships in the behavioral sciences: the fact that people choose their own environments to some extent and thus create correlations between behavioral characteristics and environmental variables. As will be illustrated extensively in the next two chapters, the only way to ensure that selection bias is not operating is to conduct a true experiment in which the key variable is manipulated.

Chapter 6

Getting Things Under Control: The Case of Clever Hans

Learning Objectives

6.1 Explain John Snow's logic in testing his hypothesis on the spread of cholera

6.2 Describe how experimental control and variable manipulation reveal the causation of behavior

This chapter starts with a quiz. Don't worry; it's not about what you read in the last chapter. In fact, it should be easy because it's about the observable motion of objects in the world, something with which we have all had much experience. There are just three questions in the quiz.

For the first, you will need a piece of paper. Imagine that a person is whirling a ball attached to a string around his or her head. Draw a circle that represents the path of the ball as viewed from above the person's head. Draw a dot somewhere on the circle and connect the dot to the center of the circle with a line. The line represents the string, and the dot represents the ball at a particular instant in time. Imagine that at exactly this instant, the string is cut. Your first task is to indicate with your pencil the subsequent flight of the ball.

For your next problem, imagine that you are a bomber pilot flying toward a target at 500 miles per hour at a height of 20,000 feet. To simplify the problem, assume that there is no air resistance. The question here is, at which location would you drop your bomb: before reaching the target, directly over the target, or when you have passed the target? Indicate a specific distance in front of the target, directly over the target, or a specific distance past the target.

Finally, imagine that you are firing a rifle from shoulder height. Assume that there is no air resistance and that the rifle is fired exactly parallel to the ground. If a bullet that is dropped from the same height as the rifle takes one-half second to hit the ground, how long will it take the bullet that is fired from the rifle to hit the ground if its initial velocity is 2,000 feet per second?

And the answers—oh, yes, the answers. They appear later on in this chapter. But, first, in order to understand what the accuracy of our knowledge about moving objects has to do with psychology, we need to explore more fully the nature of the experimental logic that scientists use. In this chapter, we will discuss principles of experimental control and manipulation.

Snow and Cholera

In his studies of pellagra, Joseph Goldberger was partially guided by his hunch that the disease was not contagious. But 70 years earlier, John Snow, in his search for the causes of cholera, bet the opposite way and also won (Johnson, 2007; Shapin, 2006). Many competing theories had been put forth to explain the repeated outbreaks of cholera in London in the 1850s. Many doctors believed that the exhalations of victims were inhaled by others who then contracted the disease. This was called the miasmal theory. By contrast, Snow hypothesized that the disease was spread by the water supply, which had become contaminated with the excrement of victims.

Snow set out to test his theory. Fortunately, there were many different sources of water supply in London, each serving different districts, so the incidence of cholera could be matched with the different water supplies, which varied in degree of contamination. Snow realized, however, that such a comparison would be subject to severe selection biases (recall the discussion in Chapter 5). The districts of London varied greatly in wealth, so any correlation between water supply and geography could just as easily be due to any of the many other economically related variables that affect health, such as diet, stress, job hazards, and quality of clothing and housing. In short, the possibility of obtaining a spurious correlation was nearly as high as in the case of the pellagra–sewage example discussed in Chapter 5. However, Snow was astute enough to notice and to exploit one particular situation that had occurred.

In one part of London, there happened to be two water companies that supplied a single neighborhood unsystematically. That is, on a particular street, a few houses were supplied by one company, then a few by the other, because in earlier days the two companies had been in competition. There were even cases in which a house had water from a company different from the one supplying the houses on either side of it. Thus, Snow had uncovered a case in which the SES of the people supplied by two water companies was virtually identical, or at least as close as it could be in a naturally occurring situation like this. Such a circumstance would still not have been of any benefit if the water from the two companies had been equally contaminated because Snow would have had no difference to associate with cholera incidence. Fortunately, this was not the case.

After the previous London cholera epidemic, one company, the Lambeth Company, had moved upstream on the Thames to escape the London sewage. The Southwark and Vauxhall Company, however, had stayed downstream. Thus, the probability was that the water of the Lambeth Company was much less contaminated than the water of the Southwark and Vauxhall Company. Snow confirmed this by chemical testing. All that remained was to calculate the cholera death rates for the houses supplied by the two water companies. The rate for the Lambeth Company was 37 deaths per 10,000 houses, compared with a rate of 315 per 10,000 houses for the Southwark and Vauxhall Company.

In this chapter, we will discuss how the Snow and Goldberger stories both illustrate the logic of scientific thinking. Without an understanding of this logic, the things scientists do may seem mysterious, odd, or downright ridiculous.

Comparison, Control, and Manipulation

Although many large volumes have been written on the subject of scientific methodology, it is simply not necessary for the layperson, who may never actually carry out an experiment, to become familiar with all the details and intricacies of experimental design. The most important characteristics of scientific thinking are actually quite easy to grasp. Scientific thinking is based on the ideas of *comparison, control*, and *manipulation.* To achieve a more fundamental understanding of a phenomenon, a scientist compares conditions in the world. Without this comparison, we are left with isolated instances of

observations, and the interpretation of these isolated observations is highly ambiguous, as we saw in Chapter 4 in our discussion of testimonials and case studies.

By comparing results obtained in different—but controlled—conditions, scientists rule out certain explanations and confirm others. The essential goal of experimental design is to *isolate a variable.* When a variable is successfully isolated, the outcome of the experiment will eliminate a number of alternative theories that may have been advanced as explanations. Scientists weed out the maximum number of incorrect explanations either by directly controlling the experimental situation or by observing the kinds of naturally occurring situations that allow them to test alternative explanations.

The latter situation was illustrated quite well in the cholera example. Snow did not simply pick any two water companies. He was aware that water companies might supply different geographic localities that had vastly different health-related socioeconomic characteristics. Merely observing the frequency of cholera in the various localities would leave many alternative explanations of any observed differences in cholera incidence. Highly cognizant that science advances by eliminating possible explanations (recall our discussion of falsifiability in Chapter 2), Snow looked for and found a comparison that would eliminate a large class of explanations based on health-related correlates of SES.

Snow was fortunate to find a naturally occurring situation that allowed him to eliminate alternative explanations. But it would be absurd for scientists to sit around waiting for circumstances like Snow's to occur. Instead, most scientists try to restructure the world in ways that will differentiate alternative hypotheses. To do this, they must manipulate the variable believed to be the cause (contamination of the water supply, in Snow's case) and observe whether a differential effect (cholera incidence) occurs while they keep all other relevant variables constant. The variable manipulated is called the *independent variable* and the variable upon which the independent variable is posited to have an effect is called the *dependent variable.*

Thus, the best experimental design is achieved when the scientist can manipulate the variable of interest and control all the other extraneous variables affecting the situation. Note that Snow did not do this. He was not able to manipulate the degree of water contamination himself but instead found a situation in which the contamination varied and in which other variables, mainly those having to do with SES, were—by lucky chance—controlled. However, this type of naturally occurring situation is not only less common but also less powerful than direct experimental manipulation.

Joseph Goldberger did directly manipulate the variables he hypothesized to be the causes of the particular phenomenon he was studying (pellagra). Although Goldberger observed and recorded variables that were correlated with pellagra, he also directly manipulated two other variables in his series of studies. Recall that he induced pellagra in a group of prisoners given a low-protein diet and also failed to induce it in a group of volunteers, including himself and his wife, who ingested the excrement of pellagra victims. Thus, Goldberger went beyond observing naturally occurring correlations and created a special set of circumstances designed to yield data that would allow a stronger inference by ruling out a wider set of alternative explanations than Snow's did. This is precisely the reason why scientists attempt to manipulate a variable and to hold all other variables constant: in order to eliminate alternative explanations.

explanations.

Random Assignment in Conjunction with Manipulation Defines the True Experiment

We are not saying here that Snow's approach was without merit. But scientists do prefer to manipulate the experimental variables more directly because direct manipulation generates stronger inferences. Consider Snow's two groups of subjects: those

whose water was supplied by the Lambeth Company and those whose water was supplied by the Southwark and Vauxhall Company. The mixed nature of the water supply system in that neighborhood probably ensured that the two groups would be of roughly equal social status. However, the drawback of the type of research design used by Snow is that the subjects themselves determined which group they would be in (self selection). They did this by signing up with one or the other of the two water companies years before. We must consider why some people signed up with one company and some with another. Did one company offer better rates? Did one advertise the medicinal properties of its water? We do not know. The critical question is, might people who respond to one or another of the advertised properties of the product differ in other, health-related ways? The answer to this question has to be, it is a possibility.

A design such as Snow's cannot rule out the possibility of spurious correlates more subtle than those that are obviously associated with SES. This is precisely the reason that scientists prefer direct manipulation of the variables they are interested in. When manipulation is combined with a procedure known as *random assignment* (in which the subjects themselves do not determine which experimental condition they will be in but, instead, are randomly assigned to one of the experimental groups), scientists can rule out alternative explanations of data patterns that depend on the particular characteristics of the subjects. Random assignment ensures that the people in the conditions compared are roughly equal on all variables because, as the sample size increases, random assignment tends to balance out chance factors. This is because the assignment of the participants is left up to an unbiased randomization device rather than the explicit choices of a human. Please note here that random *assignment* is *not* the same thing as random *sampling.* The difference will be discussed in Chapter 7.

Random assignment is a method of assigning subjects to the experimental and control groups so that each subject in the experiment has the same chance of being assigned to either of the groups. Flipping a coin is one way to decide to which group each subject will be assigned. In actual experimentation, a computer-generated table of random numbers is most often used. By using random assignment, the investigator is attempting to equate the two groups on all behavioral and biological variables prior to the investigation—even ones that the investigator has not explicitly measured or thought about.

How well random assignment works depends on the number of subjects in the experiment. As you might expect, the more the better. That is, the more subjects there are to assign to the experimental and control groups, the closer the groups will be matched on all variables prior to the manipulation of the independent variable. The use of random assignment ensures that there will be no systematic bias in how the subjects are assigned to the two groups. The groups will always be matched fairly closely on any variable, but to the extent that they are not matched, random assignment removes any bias toward either the experimental or the control group. Perhaps it will be easier to understand how random assignment eliminates the problem of systematic bias if we focus on the concept of replication: the repeating of an experiment in all of its essential features to see if the same results are obtained.

Imagine an experiment conducted by a developmental psychologist who is interested in the effect of early enrichment experiences for preschool children. Children randomly assigned to the experimental group receive the enrichment activi-

ties designed by the psychologist during their preschool day-care period. Children randomly assigned to the control group participate in more traditional playgroup activities for the same period. The dependent variable is the children's school achievement, which is measured at the end of the children's first year in school to see whether children in the experimental group have outperformed those in the control group.

An experiment like this would use random assignment to ensure that the groups start out relatively closely matched on all extraneous variables that could affect the dependent variable of school achievement. These extraneous variables are sometimes

called *confounding variables.* Some possible confounding variables are intelligence test scores and home environment. Random assignment will roughly equate the two groups on these variables. However, particularly when the number of subjects is small, there may still be some differences between the groups. For example, if after random assignment the intelligence test scores of children in the experimental group averaged 105.6 and those of children in the control group averaged 101.9 (this type of difference could occur even if random assignment has been properly used), we might worry that any difference in academic achievement in favor of the experimental group was due to the higher intelligence test scores of children in that group rather than to the enrichment program. Here is where the importance of replication comes in. Subsequent studies may again show IQ differences between the groups after random assignment, but the lack of systematic bias in the random assignment procedure ensures that the difference will not always be in favor of the experimental group. In fact, what the property of no systematic bias ensures is that, across a number of similar studies, any IQ differences will occur approximately half of the time in favor of the experimental group and half of the time in favor of the control group. In Chapter 8 we will discuss how multiple experiments such as these are used to converge on a conclusion.

Thus, there are really two strengths in the procedure of random assignment. One is that in any given experiment, as the sample size gets larger, random assignment ensures that the two groups are relatively matched on all extraneous variables. However, even in experiments where the matching is not perfect, the lack of systematic bias in random assignment allows us to be confident in any conclusions about cause—as long as the study can be replicated. This is because, across a *series* of such experiments, differences between the two groups on confounding variables will balance out.

The Importance of Control Groups

All sciences contain examples of mistaken conclusions drawn from studies that fell short of the full controls of the true experiment. Ross and Nisbett (1991) discuss the medical findings on the portacaval shunt, a treatment for cirrhosis of the liver that was popular many years ago. The studies on the treatment were assembled, and an interesting pattern was revealed. In 96.9 percent of the studies that did not contain a control group, the physicians judged the treatment to be at least moderately effective. In the studies in which there was a control group but in which random assignment to conditions was not used (thus falling short of true experimental design), 86.7 percent of the studies were judged to have shown at least moderate effectiveness. However, in the studies in which there was a control group formed by true random assignment, only 25 percent of the studies were judged to have shown at least moderate effectiveness. Thus, the effectiveness of this particular treatment—now known to be ineffective—was vastly overestimated by studies that did not employ complete experimental controls. The positive results found using less controlled procedures were due to placebo effects and/or biases resulting from nonrandom assignment. For example, selection effects (see Chapter 5) may operate to cause spurious positive effects when random assignment is not used. If the patients chosen for a treatment tend to be "good candidates" or tend to be those with vocal and supportive families, there may be differences between them and the control group irrespective of the effectiveness of the treatment.

The tendency to see the necessity of acquiring comparative information before coming to a conclusion is apparently not a natural one—which is why training in all the sciences includes methodology courses that stress the importance of constructing control groups. The "nonvividness" of the control group—the group treated just like the experimental group except for the absence of a critical factor—makes it difficult to see how essential such a group is. Psychologists have done extensive research on the tendency for people to ignore essential comparative (control group) information.

For example, in a much researched paradigm (Stanovich, 2010), subjects are shown a 2 × 2 matrix such as the one shown here that summarizes the data from an experiment.

	Improvement	No Improvement
Treatment	200	75
No Treatment	50	15

The numbers in the table represent the number of people in each cell. Specifically, 200 people received the treatment and showed improvement in the condition being treated, 75 received the treatment and showed no improvement, 50 received no treatment and showed improvement, and 15 received no treatment and showed no improvement. The subjects are asked to indicate the degree of effectiveness of the treatment. Many subjects think that the treatment in question is effective, and a considerable number of subjects think that the treatment has *substantial* effectiveness. They focus on the large number of cases (200) in the cell indicating people who received treatment and showed improvement. Secondarily, they focus on the fact that more people who received treatment showed improvement (200) than showed no improvement (75).

In fact, the particular treatment tested in this experiment is completely *ineffective.* In order to understand why the treatment is ineffective, it is necessary to concentrate on the two cells that represent the outcome for the control group (the no-treatment group). There we see that 50 of 65 subjects in the control group, or 76.9 percent, improved when they got *no* treatment. This contrasts with 200 of 275, or 72.7 percent, who improved when they received the treatment. Thus, the percentage of improvement is actually larger in the no-treatment group, an indication that this treatment is totally ineffective. The tendency to ignore the outcomes in the no-treatment cells and focus on the large number in the treatment/improvement cell seduces many people into viewing the treatment as effective. In short, it is relatively easy to draw people's attention away from the fact that the outcomes in the control condition are a critical piece of contextual information in interpreting the outcome in the treatment condition.

Many fields, not just psychology, have gradually developed an awareness of the necessity of comparative information when evaluating evidence. This is a fairly recent development even in the medical field, for instance (Gawande, 2010; Lewis, 2017). Neurologist Robert Burton (2008) describes well the path that medicine has taken: "being a good doctor requires sticking with the best medical evidence, even if it contradicts your personal experience. We need to distinguish between gut feeling and testable knowledge, between hunches and empirically tested evidence" (p. 161).

Business and governments have increasingly turned to controlled experimentation to find out how to optimize their policies, and such studies can lead to surprises. Some years ago, the State of Oregon sought to test the long-held idea that providing uninsured citizens with health insurance would drive down government health costs because insured people would be less likely to walk into emergency rooms for treatment (Sanger-Katz, 2014). The uninsured using emergency rooms is a cause of increased government and hospital costs. In order to test this idea, and see how much the savings were, the State of Oregon ran a true experiment in which they randomly chose some insured people to receive insurance and had an equal-sized group who lost

the insurance lottery serve as the control group. This type of investigation is known as a *field experiment*—where a variable is manipulated in a nonlaboratory setting. In the case of the Oregon experiment, the findings were surprising. The experimental group who received insurance did not result in fewer government costs, and they even turned out to use emergency rooms *more* than the control group! However, not all the outcomes were negative: The experimental group was found to have better mental health and quality of life.

Another example of a field experiment was a study run to see whether some progress could be made on the problem of high school students being accepted to universities but never making it to campus in the fall (Castleman, 2015; Kirp, 2017). Not surprisingly, many of these are low income students who are the first in their families to go to college. Researchers conducted a field experiment on 5,000 students the summer before college, in which the experimental group received relevant text messages like "Have you chosen your courses yet?". Of this group, 72 percent ended up enrolled in college that fall, compared with 66.4 percent of the control group. This intervention also proved to be very cost effective.

International aid organizations are likewise turning to studies with manipulated variables (true experiments) to find out "what works" (Banerjee & Duflo, 2009; Duflo & Karlan, 2016). Aid organizations often evaluate themselves, and thus end up claiming that everything they do works, which is implausible. Such an approach of course means that money will be misspent. In order to efficiently use aid money—that is, to save more lives—it is essential to make a judgment about which programs work better than others. Field experiments help with that judgment.

It is sometimes hard for the public to understand that field experiments are necessary in order to achieve something else that they want—that tax money be used efficiently, to help the most people. For example, New York City attempted an experimental test of one of its programs—Homebase—that tries to prevent people from becoming homeless (Buckley, 2010). More people are eligible (a person must be behind on rent and in danger of eviction) for this program (which includes job training, counseling, and other aid) than can be served. Thus, the city did the logical thing to test the efficacy of the program: They randomly assigned (until the money—$23 million—ran out) some people to the Homebase program and an equal number were followed up who were not included in the Homebase program. This design allowed the city to determine how many people (either just a few, or perhaps many) were saved from homelessness by this expenditure of $23 million.

Unfortunately, many citizens and groups in New York did not see it that way. They reacted emotionally to the vivid word "experiment" and objected to this controlled study that would allow the city to spend its money better. They thought the homeless were being treated like guinea pigs or lab rats. What these critics were forgetting was that no one was being *denied* service by this experiment. The *same* number of people would receive Homebase whether they were randomly assigned or not. The only difference was that by collecting information from the control group, rather than simply ignoring those who were not in the program, the city would be able to determine whether the program actually *works*!

The confusions about field experiments in the Homebase example are quite common ones. People do not seem to understand that by doing field experiments on the effects of social aid in real environments, we can maximize the number served by finding out what works best. As one international aid expert, Esther Duflo, noted, "it doesn't seem like a hugely innovative view of the world, but most people who are not economists don't get it. They don't get the idea that there are budget constraints" (Parker, 2010, p. 87). It is easy to detect a little frustration in Duflo's voice as we read this. Duflo is running up against something we will discuss many times in this book—what is obvious to a scientist is often totally missed by a layperson. It seems obvious

to Duflo that, with a fixed aid budget, the number of people served from a given program is a certain number. Another program that is more efficient would serve more people for the same fixed cost. And the only way to figure out if a program is more efficient is to run a true experiment.

Perhaps a reframing would help people. One of Duflo's colleagues helping to run the experiments on aid in impoverished countries notes that she is often told that "you shouldn't be experimenting on people" and replies "OK, so you have no idea whether your program works—and *that's* not experimental?" (p. 87). She has the right idea

in this response. The status quo—the original program being tested for its efficacy—could be called an experiment too, just a poorly designed one. The program currently being run *is* an experiment—just one without the proper controls! That is, it's a condition without a control group. It is "experimenting on people" too! This type of framing might help to dissolve the silly resistance to objective methods for finding out what helps people the most.

People seem to think, irrationally, that experiments, with their necessity of an untreated control group, somehow violate a notion of "fairness". This attitude comes about because the labeling of the control group seems to imply that some people are being "left out". This is a fallacy, but it is a powerful illusion that leads some people to irrationally oppose experiments. Psychologist Dan Ariely (2016) describes how a colleague wanted to give an incentive to half of a high school to see if it might boost attendance. The school objected to the study because it "left out" half of the school's students (the control group). After this experience, Ariely's colleague tried a different strategy. He told a school's principal that he wanted to give the students an incentive, but his budget constraints meant that he could only cover half the students. He asked the school personnel for their suggestions on how to distribute the incentive. You guessed it. They suggested it be distributed randomly—just what he would wanted to do anyway! Suddenly, when the same experiment was reframed from "the control group doesn't get anything" to "half our school will get something" the attitudes completely changed.

The Case of Clever Hans, the Wonder Horse

The necessity of eliminating alternative explanations of a phenomenon by the use of experimental control is well illustrated by a story that is famous in the annals of behavioral science: that of Clever Hans, the mathematical horse. Over 100 years ago, a German schoolteacher presented to the public a horse, Clever Hans, who supposedly knew how to solve mathematical problems. When Hans was given addition, subtraction, and multiplication problems by his trainer, he would tap out the answer to the problems with his hoof. The horse's responses were astoundingly accurate.

Many people were amazed and puzzled by Clever Hans's performance. Was the horse really demonstrating an ability thus far unknown in his species? Imagine what the public must have thought. Compelling testimonials to Hans's unique ability appeared in the German press. A group of "experts" observed Hans and attested to his abilities. Everyone was baffled. And bafflement was bound to remain as long as the phenomenon was merely observed in isolation—without controlled observations being carried out. The mystery was soon dispelled, however, when a psychologist, Oskar Pfungst, undertook systematic studies of the horse's ability (Heinzen et al., 2014).

In the best traditions of experimental design, Pfungst systematically manipulated the conditions under which the animal performed, thus creating "artificial" situations (see Chapter 7) that would allow tests of alternative explanations of the horse's performance. After much careful testing, Pfungst found that the horse did have a special ability, but it was not a mathematical one. In fact, the horse was closer to being a behavioral scientist than a mathematician. You see, Hans was a very careful observer of human behavior. As it was tapping out its answer, it would watch the

head of the trainer or other questioner. As Hans approached the answer, the trainer would involuntarily tilt his head slightly, and Hans would stop. Pfungst found that the horse was extremely sensitive to visual cues. It could detect extremely small head movements. Pfungst tested this hypothesis by having the problems presented by a trainer who was occluded from the horse's view. The animal lost its "mathematical abilities" when the trainer was out of view. (Modern versions of Pfungst's techniques are used to test whether cues from their handlers are affecting drug-sniffing police dogs, Lit et al., 2011.)

The case of Clever Hans is a good context in which to illustrate the importance of carefully distinguishing between the *description* of a phenomenon and the *explanation* of a phenomenon. That the horse tapped out the correct answers to mathematical problems presented by the trainer is not in dispute. The trainer was not lying. Many observers attested to the fact that the horse actually did tap out the correct answers to mathematical problems presented by the trainer. It is in the next step that the problem arises: making the inference that the horse was tapping out the correct answers *because* the horse had mathematical abilities. Inferring that the horse had mathematical abilities was a *hypothesized explanation* of the phenomenon. It did not follow logically—from the fact that the horse tapped out the correct answers to mathematical problems—that the horse had mathematical abilities. Positing that the horse had mathematical abilities was only one of many possible explanations of the horse's performance. It was an explanation that could be put to empirical test. When put to such a test, the explanation was falsified.

The Clever Hans case shows the danger of jumping too fast from a behavioral description to a theoretical explanation for the behavioral pattern. A contemporary example concerns the behavioral claim often heard in political discussions that "women earn only 77 cents for every dollar a man makes" (or 79 cents, the number varies). There is nothing wrong with the statement per se, because it is simply a descriptive fact. But in political discussions, an unjustified theoretical leap is often made when this descriptive fact is used to press the theoretical argument that the "77 cents" statement is a direct indicator of discrimination: that it means that women are paid less for doing the same work as men. That statement is a theoretical inference, not a descriptive fact. And the theoretical inference is wrong. Women are *not* paid 23 percent less for doing exactly the same work as men. The 77 cents figure refers to total *earnings*, not *wages* in the same job. It is possible for there to be a substantial gap in *earnings* even when the wages paid to males and females are absolutely equal for doing the same work.

Many people do not know that the 77 cents figure is arrived at simply by adding up all the income of full time (over 30 hours) workers and dividing by the number of workers. It does not take into account the *type* of occupation, the years experience in that occupation, the exact number of hours worked, prior education, overtime worked, the job classification, skills involved, and a host of other variables. When all of these variables are statistically controlled (using the multiple regression techniques mentioned in Chapter 5) the pay gap largely disappears (Bertrand et al., 2010; Black et al., 2008; CONSAD, 2009; Kolesnikova & Liu, 2011; O'Neill & O'Neill, 2012; Solberg & Laughlin, 1995). Thus, the descriptive fact of the 77 cents earnings gap is not evidence for the *theory* that women earn 23 percent less than men for doing exactly the same job. The "77 cents" statement should not be used, in political discussions, as a justification for social policies that assume the existence of large gender discrimination in *wages*.

The error in the case of the pay gap, the error in jumping too fast from descriptive facts to a particular hypothetical explanations, is just like what happened in the Clever Hans case. Before the intervention of Pfungst, the experts who looked at the horse had made this fundamental error: They had not seen that there might be alternative explanations of the horse's performance. They thought that, once they had observed that the trainer was not lying and that the horse actually did tap out the correct answers to mathematical problems, it necessarily followed that the horse had mathematical

to mathematical problems, it necessarily followed that the horse had mathematical abilities. Pfungst was thinking more scientifically and realized that that was only one of many possible explanations of the horse's performance, and that it was necessary to set up controlled conditions in order to differentiate alternative explanations. By having the horse answer questions posed by the trainer from behind a screen, Pfungst set up conditions in which he would be able to differentiate two possible explanations: that the horse had mathematical abilities or that the horse was responding to visual cues. If the horse actually had such abilities, putting the trainer behind a screen should make no difference in its performance. On the other hand, if the horse was responding

to visual cues, then putting the trainer behind a screen should disrupt its performance. When the latter happened, Pfungst was able to rule out the hypothesis that the horse had mathematical abilities (Heinzen et al., 2014).

Note also the link here to the principle of parsimony discussed in Chapter 3—the principle that states that when two theories have the same explanatory power, the simpler theory (the one involving fewer concepts and conceptual relationships) is preferred. The two theories in contention here—that the horse had true mathematical abilities and that the horse was reading behavioral cues—are vastly different in parsimony. The latter requires no radical adjustments in prior psychological and brain theory. It simply requires us to adjust slightly our view of the potential sensitivity of horses to behavioral cues (which was already known to be high). The former theory—that horses can truly learn arithmetic—requires us to alter dozens of concepts in evolutionary science, cognitive science, comparative psychology, and brain science. It is unparsimonious in the extreme because it does not cohere with the rest of science and, thus, requires that many other concepts in science be altered if it is to be considered true (we shall discuss the so-called principle of connectivity in Chapter 8).

Clever Hans in the 1990s and in the Present Day

The Clever Hans story is a historical example that has been used in methodology classes for many years to teach the important principle of the necessity of experimental control. No one ever thought that an *actual* Clever Hans case could happen *again*—but it did. Throughout the early 1990s, researchers the world over watched in horrified anticipation—almost as if observing cars crash in slow motion—while a modern Clever Hans case unfolded before their eyes and had tragic consequences.

Autism is a developmental disability characterized by impairment in reciprocal social interaction, delayed and often qualitatively abnormal language development, and a restricted repertoire of activities and interests (Baron-Cohen, 2008). The extremely noncommunicative nature of many autistic children, who may be normal in physical appearance, makes the disorder a particularly difficult one for parents to accept. It is, therefore, not hard to imagine the excitement of parents of autistic children when, in the late 1980s and early 1990s, they heard of a technique coming out of Australia that enabled autistic children who had previously been totally nonverbal to communicate. This technique for unlocking communicative capacity in nonverbal autistic individuals was called facilitated communication, and it was uncritically trumpeted in the most popular media outlets of that era (Hagen, 2012; Heinzen et al., 2014; Offit, 2008). The claim was made that autistic individuals and other children with developmental disabilities who had previously been nonverbal had typed highly literate messages on a keyboard when their hands and arms had been supported over the typewriter by a sympathetic "facilitator." Not surprisingly, these startlingly verbal performances on the part of autistic children who had previously shown very limited linguistic behavior spawned incredible hopes among frustrated parents of autistic children. It was claimed that the technique also worked for nonverbal individuals with severe intellectual disability.

Although the excitement of the parents is easy to understand, the gullibility of many professionals is not so easy to accept. Unfortunately, claims for the efficacy of

many professionals is not so easy to accept. Unfortunately, claims for the efficacy of facilitated communication were disseminated to hopeful parents by many media outlets before any controlled studies had been conducted. Had the professionals involved had minimal training in the principles of experimental control, they should have immediately recognized the parallel to the Clever Hans case. The facilitator, almost always a sympathetic individual who was genuinely concerned that the child succeed, had numerous opportunities to consciously or unconsciously direct the child's hand to the vicinity of keys on the keyboard. That cuing by the facilitator was occurring should also have been suggested by the additional observation that the children

sometimes typed out complicated messages while not even looking at the keyboard. Additionally, highly literate poetic English prose was produced by children who had not been exposed to the alphabet. For example, one child allegedly typed "Am I a slave or am I free? Am I trapped or can I be seen as an easy and rational spirit?" (Offit, 2008, p. 7). Another won an international writing competition (Hagen, 2012).

A number of controlled studies have been reported that have tested the claims of facilitated communication by using appropriate experimental controls. Each study has unequivocally demonstrated the same thing: The autistic child's performance depended on tactile cuing by the facilitator (Hagen, 2012; Heinzen et al., 2014; Offit, 2008). The controls used in several of the studies resembled those of the classic Clever Hans case. A controlled situation was set up in which both the child and the facilitator were presented with a drawing of an object but in which they could not see each other's drawing. When both child and facilitator were looking at the *same* drawing, the child typed the correct name of the drawing. However, when the child and the facilitator were shown *different* drawings, the child typed the name of the facilitator's drawing, not the one at which the child was looking. Thus, the responses were determined by the facilitator rather than the child.

The conclusion that facilitated communication was a Clever Hans phenomenon and not a breakthrough therapeutic technique brought no joy to the investigators involved in conducting the studies. But this sad story gets even worse. At some centers, during facilitated sessions on the keyboard, clients reported having been sexually abused by a parent in the past (Offit, 2008). Children were removed from their parents' homes, only to be returned when the charges of abuse proved to be groundless.

As a result of the controlled studies, competent professional opinion finally began to be heard above the media din. Importantly, it is increasingly recognized that treatments that lack empirical foundation are not benignly neutral ("Oh, well, it might work, and so what if it doesn't?"). The implementation of unproven treatments has real costs. Also, with facilitated communication, we have another example of the harm done by reliance on testimonial evidence and the fallacy of the idea that therapeutic fads and pseudoscience do no harm (see Chapter 4). We can also see that there is simply no substitute for the control and manipulation of the experimental method when we want to explain behavior.

However, decades after facilitated communication was debunked as a bogus technique, it still reappears in schools and in popular culture. Writing about the technique making a comeback in schools, writer Kendrick Frazier (2015) called it "returned from the dead" (sometimes under new names like "supported typing"). In 2011, a few unfortunate parents were still being accused of sexual abuse by their children when the children were subjected to facilitated communication sessions (Hagen, 2012). In 2015, a Rutgers University professor thought that a client that she was "facilitating" actually consented to sexual activity. At her later trial, her defense was that she thought that he had no intellectual impairment at all because of what he typed (with her assistance) during the sessions (Radford, 2016a). CNN, MSNBC, and the BBC ran stories of cases entirely without skepticism long after the techniques was exposed as a case of the Clever Hans Syndrome (Hagen, 2012). A film touting the technique played in over 100 cities in 2011, almost 20 years after it was first debunked (Hagen, 2012). It is indeed the

bogus remedy that just will not die. On World Autism Day, April 2, 2016, Apple got in the act by introducing a video of an autistic child claimed to be writing on an iPad with the aid of facilitated communication—called "rapid prompting" in this case (Shermer, 2016: Vyse, 2016b).

Note also the link to the principle of parsimony. That the severe linguistic difficulties of autistic children could be solved by a single "magic bullet" (see Chapter 9) intervention flies in the face of decades of work on the cognitive, neuropsychological, and brain characteristics of autistic children (Baron-Cohen, 2005; Oberman & Ramachandran, 2007; Rajendran & Mitchell, 2007; Wellman et al., 2011). It would

require that too much else that we know about cognition and neurology be overturned. The existence of facilitated communication would show no connectivity with the rest of science (see Chapter 8).

Finally, the example of facilitated communication illustrates something discussed previously in the Clever Hans case: the importance of carefully distinguishing between the *description* of a phenomenon and the *explanation* of a phenomenon. The term "facilitated communication" is not a neutral description of what occurred between facilitator and child. Instead it posits a theoretical outcome—that communication actually occurred and had been truly enhanced by the facilitator. But that is the very thing that had to be proved! What we had here was a child tapping keys. Perhaps things would have proceeded more rationally had it originally been labeled "surprising tapping." What needed to be determined was whether the "surprising tapping" was true communication. The premature labeling of the phenomenon (key tapping) with a theory (that it represented true communication) likely made it more difficult for these practitioners to realize that further investigation was necessary to see if this theoretical label was warranted.

Other fields—not just psychology—struggle with the problem of prematurely labeling a phenomenon with a theory. The legal system still uses the term "shaken baby syndrome" when in fact the American Academy of Pediatrics has recommended that that term be discarded. The problem is exactly like the Clever Hans and facilitated communication examples we have been discussing. The term "shaken baby syndrome" is a *theory* of why a particular child has presented with head trauma. The *phenomenon* is the nature of the head trauma itself. The precise reason for the head trauma is what has to be *explained* by whatever theory we have of how the trauma occurred. The legal system is still working through the implications of this change in terminology that had once been standard, but that we now know to be misleading (Tuerkheimer, 2010).

Traffic safety engineers likewise feel that the term traffic "accident" carries too much theory with it (Richtel, 2016). The word accident implies randomness and unpredictability and luck—pure happenstance. Safety engineers know all too well that automobile crash risk has strong statistical relationships to many behaviors, none of which are random or happenstance. The engineers have in mind cases like St. Louis Cardinals pitcher Josh Hancock who slammed his rented SUV into a truck stopped on the highway with lights flashing (Vanderbilt, 2008). Calling the crash random and unpredictable (an "accident") seems not at all right when we consider that Hancock was speeding (a strong risk factor), had an alcohol concentration twice the legal limit (a strong risk factor), and was on a cell phone at the time of the crash (a strong risk factor). Oh, and he had crashed another SUV just two days before (Vanderbilt, 2008). Terming this an "accident" conveys a theory of randomness and unpredictability that does not seem right when the chosen behaviors were so wantonly reckless as in this case. The *description* of what happened is—a crash. As a *theory*, accident seems not quite right.

Prying Variables Apart: Special Conditions

The Goldberger pellagra example illustrates a very important lesson that can greatly

aid in dispelling some misconceptions about the scientific process, particularly as it is applied in psychology. The occurrence of any event in the world is often correlated with many other factors. In order to separate, to pry apart, the causal influence of many simultaneously occurring events, we must create situations that will never occur in the ordinary world. Scientific experimentation breaks apart the natural correlations in the world to isolate the influence of a single variable.

Psychologists operate in exactly the same manner: by isolating variables via manipulation and control. For example, cognitive psychologists interested in the reading process have studied the factors that make word perception easier or more difficult.

Not surprisingly, they have found that longer words are more difficult to recognize than shorter words. At first glance, we might think that the effect of word length would be easy to measure: Simply create two sets of words, one long and one short, and measure the difference in reader recognition speed between the two. Unfortunately, it is not that easy. Long words also tend to be less frequent in language, and frequency *itself* also affects perception. Thus, any difference between long and short words may be due to length, frequency, or a combination of these two effects. In order to see whether word length affects perception independently of frequency, researchers must construct special word sets in which length and frequency do not vary together.

Similarly, Goldberger was able to make a strong inference about causation because he set up a special set of conditions that does not occur naturally. (Considering that one manipulation involved the ingestion of bodily discharges, this is putting it mildly!) Recall that Oskar Pfungst had to set up some special conditions for testing Clever Hans, including trials in which the questioner did not know the answer. Dozens of people who merely observed the horse answer questions under normal conditions (in which the questioner knew the answer) never detected how the horse was accomplishing its feat. Instead, they came to the erroneous conclusion that the horse had true mathematical knowledge.

Likewise, note the unusual conditions that were necessary to test the claims of facilitated communication. The stimuli presented to the facilitator and the child had to be separated in a way that neither could see the stimulus presented to the other. Such unusual conditions are necessary in order to test the alternative hypotheses for the phenomenon.

Many classic experiments in psychology involve this logic of prying apart the natural relationships that exist in the world so that it can be determined which variable is the dominant cause. Psychologist Harry Harlow's famous experiments (Harlow & Suomi, 1970; Tavris, 2014) provide a case in point. Harlow wanted to test a prevailing hypothesis about infant–mother attachment: That attachment resulted from the mother providing the infant's source of food. However, the problem was that, of course, mothers provide much more than nourishment (comfort, warmth, caressing, stimulation, etc.). Harlow examined the behavior of infant macaque monkeys in situations in which he isolated only one of the variables associated with attachment by giving the animals choices among "artificial" mothers. For example, he found that the contact comfort provided by a "mother" made of terrycloth was preferred to that provided by a "mother" made of wire mesh. After two weeks of age, the infant preferred a cold terrycloth mother to a warm wire one, a finding indicating that the contact comfort was more attractive than warmth. Finally, Harlow found that the infants preferred the terrycloth mother even when their nourishment came exclusively from a wire mother. Thus, the hypothesis that attachment was due solely to the nourishment provided by mothers was falsified. This was possible only because Harlow was able to pry apart variables that naturally covary in the real world.

Creating special conditions to test for actual causal relationships is a key tool we can use to prevent pseudoscientific beliefs from attacking us like a virus (Stanovich, 2004, 2009, 2011). Consider the case of therapeutic touch (TT)—a fad that swept the North American nursing profession in the 1990s. TT practitioners massage not the

patient's body but instead the patient's so-called energy field. That is, they move their hands over the patient's body but do not actually massage it. Practitioners reported "feeling" these energy fields. Well, you guessed it. This ability to feel "energy fields" is tested properly by creating exactly the type of special conditions as in the Clever Hans and facilitated communication claims—that is, testing whether practitioners, when visually blinded, could still feel whether their hands were in proximity to a human body. Research has demonstrated the same thing as in the Clever Hans and facilitated communication cases—when vision is occluded, this ability to feel at a distance is no greater than chance (Hines, 2003; Shermer, 2005). This example actually illustrates

something that was mentioned in an earlier chapter—that the logic of the true experiment is really so straightforward that a child could understand it. This is because one of the published experiments showing that TT is ineffective was done as a school science project (Dacey, 2008).

In short, it is often necessary for scientists to create special conditions that will test a particular theory about a phenomenon. Merely observing the event in its natural state is rarely sufficient. People observed falling and moving objects for centuries without arriving at accurate principles and laws about motion and gravity. Truly explanatory laws of motion were not derived until Galileo and other scientists set up some rather artificial conditions for the observation of the behavior of moving objects. In Galileo's time, smooth bronze balls were rarely seen rolling down smooth inclined planes. Lots of motion occurred in the world, but it was rarely of this type. However, it was just such an unnatural situation, and others like it, that led to our first truly explanatory laws of motion and gravity. Speaking of laws of motion, didn't you take a little quiz at the beginning of this chapter?

Intuitive Physics

Actually, the three questions posed at the beginning of this chapter were derived from the work that psychologists have done on so-called "intuitive physics," that is, people's beliefs about the motion of objects. Interestingly, these beliefs are often at striking variance from how moving objects actually behave (Bloom & Weisberg, 2007; Riener et al., 2005).

For example, in the first problem, once the string on the circling ball is cut, the ball will fly in a straight line at a 90-degree angle to the string (tangent to the circle). McCloskey (1983) found that one-third of the college students who were given this problem thought, incorrectly, that the ball would fly in a curved trajectory. About half of McCloskey's subjects, when given problems similar to the bomber pilot example, thought that the bomb should be dropped directly over the target, thus displaying a lack of understanding of the role of an object's initial motion in determining its trajectory. The bomb should actually be dropped *five miles* before the plane reaches the target. The subjects' errors were not caused by the imaginary nature of the problem. When subjects were asked to walk across a room and, while moving, drop a golf ball on a target on the floor, the performance of more than half of them indicated that they did not know that the ball would move forward as it fell. Finally, many people are not aware that a bullet fired from a rifle will hit the ground at the same time as a bullet dropped from the same height.

You can assess your own performance on this little quiz. Chances are that you missed at least one if you have not had a physics course recently. "Physics course!" you might protest. "Of course I haven't had a physics class recently. This quiz is unfair!" But hold on a second. Why should you *need* a physics course? You have seen literally hundreds of falling objects in your lifetime. You have seen them fall under *naturally occurring* conditions. Moving objects surround you every day, and you are seeing them in their "real-life" state. You certainly cannot claim that you have not experienced moving and falling objects. Granted, you have never seen anything quite like the bul-

let example. But most of us have seen children let go of whirling objects, and many of us have seen objects fall out of planes. And besides, it seems a little lame to protest that you have not seen these exact situations. Given your years of experience with moving and falling objects, why can't you accurately predict what will happen in a situation only slightly out of the ordinary?

It is critical to understand that the layperson's beliefs are inaccurate precisely *because* his or her observations are "natural," rather than controlled in the manner of the scientist's. Thus, if you missed a question on the little quiz at the beginning of the chapter, don't feel ignorant or inadequate. Simply remember that some of the world's

greatest minds observed falling objects for centuries without formulating a physics of motion any more accurate than that of the modern high school sophomore.

Psychological research on intuitive physics demonstrates something of fundamental importance in understanding why scientists behave as they do. Despite extensive experience with moving and falling objects, people's intuitive theories of motion are remarkably inaccurate. Experience provides no inoculation against these intuitive errors. For example, experienced taxi drivers make most of the same speed and journey time errors that nonprofessional divers do (Peer & Solomon, 2012).

Intuitive Psychology

If our intuitive (or "folk") theories about objects in motion are inaccurate, it is hard to believe that our folk theories in the more complex domain of human behavior will be exceedingly accurate. Indeed, this research literature serves to warn us that personal experience is no guarantee against incorrect beliefs about human psychology. Psychologist Dan Ariely (2015) tells the story of suffering burns over 70 percent of his body as the result of an accident when he was 18 years old. He describes many months of subsequent treatment in which bandages that were removed quickly caused him great pain. The theory held by the nurses was that a quick removal (which caused a sharp pain) was preferable to slow removal which would cause a longer—although less intense—pain. After leaving the hospital and beginning his career as a psychology student, Ariely conducted experiments to test the nurses' belief. To his surprise, Ariely found that the slower procedure—lower pain intensity over a longer period—would have reduced the pain perception in such situations. He said that by the time he had finished, he realized that the nurses in the burn unit were kind and generous individuals with a lot of experience in soaking and removing bandages, but that "despite all their experience, they erred in treating the patients they cared so much about. They still didn't have the right theory about what would minimize their patients' pain. How could they be so wrong, I wondered, considering their vast experience? Perhaps other professionals might also be misunderstanding the consequences of their behaviors and make poor decisions" (p. C3, Ariely, 2015). Research indicates that intuitive judgments of pain intensity in other people are quite bad, even among physicians with much clinical experience (Tait et al., 2009).

As discussed in Chapter 4, reliance on testimonials, case study evidence, and "common practice" can often obscure the need for a control group to check the veracity of a conclusion derived from informal observation. For example, Dingfelder (2006) describes how many medical professionals believe that they should not advise individuals with Tourette syndrome (described in Chapter 2) to suppress their tics (involuntary vocal expressions). The physicians believed that this caused a so-called rebound effect—a higher rate of tics occurring after the suppression. This belief, though, is based on informal observation rather than controlled experimentation. When the proper experimentation was done—observing the number of tics systematically by comparing a period of suppression to a period of nonsuppression—it appeared that there was no "rebound" effect at all following tic suppression.

In Chapter 1, we illustrated that a number of commonsense (or folk) beliefs about

human behavior are wrong, and this was just a small sample. For example, it turns out that there is no strong evidence indicating that highly religious people are more altruistic than less religious people (Paloutzian & Park, 2005). Studies have indicated that there is no simple relationship between degree of religiosity and the tendency to engage in charitable acts, to aid other people in distress, or to abstain from cheating other people.

Incorrect intuitive theories are not limited to psychology. For example, they are rampant in the world of sport and physical fitness. For example, quantitative analyses have indicated that in football (at all levels, from high school to professional)

most coaches increase their probability of winning by going for it on fourth down when their teams are at midfield (Moskowitz & Wertheim, 2011). Similar analyses have shown that, overall, coaches should punt less and on-side kick more. Statistics prove that if coaches reoriented their strategies in these respects, they would win more games (Moskowitz & Wertheim, 2011). Now, coaches might have a variety of reasons for ignoring this statistical advice (fear of being second-guessed, for example), but these reasons do not apply to the fans. Nevertheless, fans have the incorrect intuitive theory that the coaches are right.

Incorrect beliefs about human behavior can have very practical consequences. Keith and Beins (2008) mention that among their students, typical views about cell phones and driving are captured by statements such as “Talking doesn’t impair my driving” and “I talk on the phone to keep myself from falling asleep.” The students seem completely oblivious to the fact that driving while using a cell phone (even a hands-free phone) seriously impairs concentration and attention (Kunar et al., 2008; Richtel, 2014; Strayer et al., 2016; Strayer & Drews, 2007) and is a cause of accidents and deaths (McEvoy et al., 2005; Novotny, 2009; Parker-Pope, 2009; Richtel, 2014). It is just as dangerous as drunk driving. Texting while driving is particularly lethal.

The list of popular beliefs that are incorrect is long. For example, many people believe that a full moon affects human behavior. It doesn’t (Univ. of California, 2013; Foster & Roenneberg, 2008). Some people believe that “opposites attract.” They don’t (Youyou et al., 2017). Some people believe that you shouldn’t change an answer on a multiple choice test. They’re wrong (Kruger et al., 2005). Some people believe that “familiarity breeds contempt.” It doesn’t (Claypool et al., 2008; Zebrowitz et al., 2008). Some people believe that people behave like robots under hypnosis. They don’t (Lilienfeld, 2014). And the list goes on and on and on (see Lilienfeld et al., 2010).

The many inadequacies in people’s intuitive theories of behavior illustrate why we need the controlled experimentation of psychology: so that we can progress beyond our flat-earth conceptions of human behavior to a more accurate scientific conceptualization.

Summary

The heart of the experimental method involves manipulation and control. This is why an experiment allows stronger causal inferences than a correlational study. In a correlational study, the investigator simply observes whether the natural fluctuation in two variables displays a relationship. By contrast, in a true experiment the investigator manipulates the variable hypothesized to be the cause and looks for an effect on the variable hypothesized to be the effect while holding all other variables constant by control and randomization. This method removes the third-variable problem present in correlational studies. The third-variable problem arises because, in the natural world, many different things are related. The experimental method may be viewed as a way of prying apart these naturally occurring relationships. It does so because it isolates one particular variable (the hypothesized cause) by manipulating it and holding everything else constant. However, in order to pry apart naturally occurring relationships, scientists often have to create special conditions that are unknown in the natural world.

Chapter 7

"But It's Not Real Life!": The "Artificiality" Criticism and Psychology

Learning Objectives

7.1 Explain how the purpose of the experiment determines its design

7.2 Summarize the applicability of theory in areas of psychological research

Having covered the basics of experimental logic in the previous two chapters, we are now in a position to consider some often-heard criticisms of the field of psychology. In particular, we will discuss at length the criticism that psychology experiments are useless because they are artificial and not like "real life."

Why Natural Isn't Always Necessary

From the discussion in Chapter 6, it should already be fairly clear why this criticism is invalid. As was illustrated in that chapter, the artificiality of scientific experimentation is not a weakness but actually the very thing that gives the scientific method its unique power to yield explanations about the nature of the world. Contrary to common belief, the artificiality of scientific experiments is not an accidental oversight. It is intentionally sought (it's a feature, not a bug!). Scientists *deliberately* set up conditions that are unlike those that occur naturally because this is the only way to separate the many inherently correlated variables that determine events in the world. To use a phrase from Chapter 6, scientists set up special conditions in order to *pry variables apart*.

Sometimes the necessary conditions already exist naturally, as in the example

Sometimes the necessary conditions already exist naturally, as in the example of Snow and cholera. More often, this is not the case. The scientist must manipulate events in new and sometimes strange ways, as in the example of Goldberger and pellagra. In many instances, these manipulations cannot be accomplished in natural environments, and the scientist finds it necessary to bring the phenomenon into the laboratory, where more precise control is possible.

Indeed, some phenomena would be completely impossible to discover if scientists were restricted to observing "natural" conditions. Physicists probing the most fundamental characteristics of matter build gigantic mile-long accelerators that induce collisions between elementary particles. Some of the by-products of these collisions

are new particles that exist for less than a billionth of a second. The properties of these new particles, however, have implications for theories of atomic structure. Many of these new particles would not ordinarily exist on earth, and even if they did, there certainly would be no chance of observing them naturally. Yet few people doubt that this is how physicists should conduct their research—that probing nature in unusual and sometimes bizarre ways is a legitimate means of coming to a deeper understanding of the universe. Somehow, though, practices that seem reasonable for physicists are often viewed as invalid when used by psychologists.

Many psychologists who have presented experimental evidence on behavior to an audience of laypersons have heard the lament "But it's not real life!" This remark reflects the belief that studying human psychology in the laboratory is somehow strange. This objection also contains the assumption that knowledge cannot be obtained unless natural conditions are studied.

It is not commonly recognized that many of the techniques used by the psychologist that are viewed as strange by the public are in no way unique to psychology; instead, they are manifestations of the scientific method as applied to behavior. Restriction to real-life situations would prevent us from discovering many things. For example, biofeedback techniques are now used in a variety of areas such as migraine and tension headache control, hypertension treatment, and relaxation training (deCharms et al., 2005; Maizels, 2005). These techniques developed out of research indicating that humans could learn partial control of their internal physiological processes if they could monitor the ongoing processes via visual or auditory feedback. Of course, because humans are not equipped to monitor their physiological functions via external feedback, the ability to control such processes does not become apparent *except* under special conditions.

Consider saccadic eye movements, often a focus of researchers studying the reading process (Seidenberg, 2017). People have the impression that their eyes move smoothly across the page when reading, but actually they don't. Introspection is not a very good guide to the reading process. The eyes are stationary most of the time during reading and actually move during brief, so-called saccadic movements of 20–40 milliseconds. In between saccades, the eyes are relatively stationary for 200–300 millisecond periods. In short, contrary to introspection, your eyes are stationary most of the time during reading, but they also jump 3–4 times per second. During the jumps you are functionally blind by the way! None of these facts can be discerned in detail without special conditions and special instrumentation.

The Random Sample Versus Random Assignment Confusion

Sometimes, however, the "it's not real life" complaint arises from a different type of confusion about the purposes of psychological experimentation, one that is actually quite understandable. Through media exposure, many people are familiar with survey research, particularly in the form of election and public opinion polling. There is now a

growing awareness of some of the important characteristics of election polling. In particular, the media have given more attention to the importance of a random, or representative, sample for the accuracy of public opinion polls. This attention has led many people to believe, mistakenly, that random samples and representative conditions are an essential requirement of all psychological investigations. Because psychological research seldom uses random samples of subjects, the application of the random sample criterion by the layperson seems to undermine most psychological investigations and to reinforce the criticism that the research is invalid because it doesn't reflect real life. Actually, it is *not* necessary for every psychological investigation to employ a random sample of participants.

Random sampling and random assignment (discussed in Chapter 6) are not the same thing. Because they both have the term "random" in them, many people come to think that random assignment and random sampling are the same. Actually they are very different concepts—similar only in that they make use of the properties of random number generation. But they are used for very different purposes.

Random sampling refers to how subjects are chosen to be part of a study. As noted previously, random sampling is not a requirement for all research, but when it does become necessary (in survey research, consumer research, or election polling, for example), it refers to drawing a sample from the population in a manner that ensures that each member of the population has an equal chance of being chosen for the sample. The sample that is drawn then becomes the subject of the investigation. It is important to understand that the investigation could be either a correlational study or a true experiment. It is not a true experiment unless random assignment is *also* used.

Random assignment is a requirement of a true experiment in which an experimental group and a control group are formed by the experimenter. Random assignment is achieved when each subject is just as likely to be assigned to the control group as to the experimental group. This is why a randomizing device such as a coin flip (or more often, a specially prepared table of random numbers) is employed—because it displays no bias in assigning the subjects to groups.

The best way to keep in mind that random assignment and random sampling are not the same thing is always to be clear that any of the four combinations can occur: nonrandom sampling without random assignment, nonrandom sampling with random assignment, random sampling without random assignment, and random sampling with random assignment. Most psychological research does not employ random sampling. The research involves theory testing, as we will see in the next section, and a convenience sample is all that is necessary. If random assignment is employed in the study, then it becomes a true experiment. If random assignment is not employed, then the study is a correlational investigation. Many studies that do use random sampling do not employ random assignment because they are surveys and are only looking for associations—that is, they are correlational investigations.

Theory-Driven Research Versus Direct Applications

In some kinds of applied research, the goal is to relate the results of the study directly to a particular situation. Election polling is an example of directly applied research. The goal is to predict a specific behavior in a very specific setting—in this case, voting on election day. Here, where the nature of the application is direct, questions of the randomness of the sample and the representativeness of the conditions are important because the findings of the study are going to be applied directly.

However, it would be a mistake to view this class of research as typical. The vast majority of research studies in psychology (or any other science, for that matter) are conducted with a very different purpose in mind. Their purpose is to advance theory. The findings of most research are applied only *indirectly* through modifications in a

theory that, in conjunction with other scientific laws, is then applied to some practical problem. In short, most theory-driven research seeks to test theories of psychological processes rather than to generalize the findings to a particular real-world situation.

Research that focuses primarily on theory testing is often termed *basic research*. Whereas in applied research the purpose of the investigation is to go from data directly to a real-world application, basic research focuses on theory testing. However, it is probably a mistake to view the basic-versus-applied distinction solely in terms of whether a study has practical applications, because this difference often simply boils down to a matter of time. Applied findings are of use immediately. Basic research

often gets applied at a much later time, and often after many twists and turns in the evolution of knowledge.

The history of science is filled with examples of theories or findings that eventually solved a host of real-world problems even though the scientists who developed the theories and/or findings did not intend to solve a specific practical problem. For example, a group of researchers at the University of Texas Southwestern Medical Center was seeking to genetically engineer a population of rats with arthritis in order to study that inflammatory disease. Unexpectedly, their rats also developed inflammation of the intestines (Fackelman, 1996) similar to ulcerative colitis. The scientists now had an animal model of the human disease. Whether these scientists make any progress on arthritis (their original problem), it now looks as if they have made a substantial contribution to the eventual treatment of ulcerative colitis and Crohn's disease.

Such indirect connections are common in science. The drug company Pfizer was looking for a new heart treatment when it discovered Viagra (Gladwell, 2010). Developments in the abstract field of number theory led to the encryption technologies that made e-commerce possible (Reif, 2016).

Psychologist Walter Mischel's (2015) famous "marshmallow study" is another example of early basic research leading to a range of practical applications. His procedure involved telling four-year-old children that they will receive a small reward (one marshmallow) or a larger reward (two marshmallows). The child gets the larger reward if, after the experimenter leaves the room, the child waits until the experimenter returns and does not recall the experimenter by ringing a bell. If the bell is rung before the experimenter returns, the child will get only the smaller reward. The dependent variable is the amount of time that the child waits before ringing the bell. His first studies using this famous delay of gratification paradigm were denied government research funding (Sleek, 2015), and he was told to seek funding from a candy company! But longitudinal studies have shown that the test, when give to four-year-olds, predicts adult success. The ability to delay gratification in childhood predicts such important life outcomes as drug use, obesity levels, and SAT scores. Mischel's (2015) work has been used in several important programs to develop children's self control skills (Winerman, 2014).

Thus, we must recognize that, although some research is designed to predict events directly in a specific environmental situation, much scientific research is basic research designed to test theory. Researchers who conduct applied and basic research have completely different answers to the question, how do these findings apply to real life? The former answers, "Directly, provided that there is a reasonably close relationship between the experimental situation and the one to which the findings are to be applied." Thus, questions of the random sampling of subjects and the representativeness of the experimental situation are relevant to the applicability of the results. However, the investigator in a theory-testing study answers that his or her findings do not apply *directly* to real life, and that the reason for conducting the study is *not* to produce findings that would be applicable to some specific environmental situation. Therefore, this scientist is not concerned with questions of how similar the subjects of

the study are to some other group or whether the experimental situation mirrors some real-life environment. Does this mean, then, that these findings have no implications for the real world? No. These findings apply directly not to a particular situation but to a theory. The theory may, at some later date, in conjunction with other scientific laws, be applied to a particular problem.

This type of indirect application through theory has become quite common in some areas of psychology. For example, years ago when cell phones were first introduced, many cognitive psychologists immediately began to worry about the implications for safety when people began to use them while driving automobiles. The psychologists

immediately expected that cell phone use would cause additional accidents—and not just because the phone would take a hand off the wheel. Instead, what they were worried about was the attentional requirements of talking on the cell phone. What is important to realize is that the psychologists became worried about cell phone use in cars long before there was a single experimental study of actual cell phone use and its relation to accidents (Strayer et al., 2016). The psychologists made their prediction of accident problems with cell phones through *theory*, in this case theories of limited-capacity attention that were decades old (e.g., Kahneman, 1973). Cell phone use in a car clearly falls within the domain of those theories, which had been established through voluminous experimentation (literally hundreds of laboratory studies). When in fact the actual studies of real cell phone use were done, they confirmed the prediction derived from psychological theories of attention: Cell phone use is indeed a cause of motor vehicle accidents—and hands-free phones do not solve the attentional problem, which is the main cause of the accidents (Insurance Institute for Highway Safety, 2005; Kunar et al., 2008; Levy et al., 2006; McEvoy et al., 2005; Richtel, 2014; Strayer & Drews, 2007; Strayer et al., 2016).

Applications of Psychological Theory

We have described how the purpose of most research is to develop theory rather than to predict events in a specific environment. And we have described how the findings of most research are applied indirectly, through theory, rather than directly in a specific environmental situation. Given these facts, though, it is legitimate to ask how much application through theory has been accomplished in psychology. That is, have psychology's theories been put to this test of generality?

On this point, we must admit that the record is mixed. But it is wise to keep psychology's diversity in mind here. It is true that some areas of research have made only modest progress along these lines. However, other areas have quite impressive records of experimentally derived principles of considerable explanatory and predictive power.

Consider the basic behavioral principles of classical and operant conditioning. These principles and their elaborating laws were developed almost entirely from experimentation on nonhuman subjects, such as pigeons and rats, in highly artificial laboratory settings. Yet these principles have been successfully applied to a wide variety of human problems, including the treatment of autistic children, the treatment of alcoholism and obesity, the management of residents in psychiatric hospitals, insomnia interventions, and the treatment of phobias, to name just a few.

The principles from which these applications were derived were identified precisely because the laboratory experimentation allowed researchers to specify the relationships between environmental stimuli and behavior with an accuracy not possible in a natural situation, in which many behavioral relationships may operate simultaneously. As for the use of nonhuman subjects, in many cases, theories and laws derived from their performance have provided good first approximations to human behavior

(Vazire & Gosling, 2003). When humans were examined, their behavior often followed laws that were very similar to those derived from other animals. This should hardly surprise anyone today, when just about every medical advance in the treatment of human illness has involved data from animal studies. For example, research with animals has contributed to developments in behavioral medicine, stress reduction, psychotherapy, rehabilitation of injured and handicapped individuals, studying the effects of aging on memory, methods to help people overcome neuromuscular disorders, understanding drug effects on fetal development, traffic safety, and the treatment of chronic pain (Gosling, 2001; Kalat, 2007; Zimbardo, 2004). Research with monkeys has led to some real advances in understanding the underlying basis of phobias and anxiety disorders

(Mineka & Zinbarg, 2006). Nevertheless, scientists, including psychologists, studying animals have come under increasing attack from animal advocates, and sometimes those attacks have been violent. J. David Jentsch, who works in the psychology department at UCLA and studies the brain circuits involved in drug addiction, had his car bombed by animal activists in 2009 and the resulting fire almost burned down his home (Collier, 2014).

Psychologists studying perceptual processes have made impressive theoretical progress, and the laws and theories they have derived have been applied to problems as diverse as radar monitoring, street lighting, and airplane cockpit design (Durso et al., 2007; Wickens et al., 2012). Much is now known about the cognitive effects of aging (Salthouse, 2012), and this new knowledge has direct implications for efforts to design systems that will help people to compensate for cognitive loss (Schaie & Willis, 2010).

Psychological studies of judgment and decision making have had implications for medical decision making, educational decision making, and economic decision making (Croskerry, 2013; Stanovich et al., 2016; Tetlock & Gardner, 2015; Thaler, 2015). The famous obedience to authority studies of Stanley Milgram were used in officer training schools of the military (Blass, 2004; Cohen, 2008). An exciting new development is the increasing involvement of cognitive psychologists in the legal system, in which problems of memory in information collection, evidence evaluation, and decision making present opportunities to test the applicability of cognitive theories (Wells et al., 2015; Wixted et al., 2015). In recent decades, theory and practice in the teaching of reading have been affected by research in cognitive psychology (Seidenberg, 2017; Willingham, 2017).

In short, psychology has been applied to “real life” in a large number of ways, but little of this is known to the public. Research psychologists have found ways of getting people to save more for their retirement and to increase their organ donations (Thaler, 2015), discovered how to influence people to get their flu shots (Price, 2009), invented behavioral programs that would reduce energy use (Attari et al., 2010), discovered ways to facilitate onscreen reading (Chamberlin, 2010), found how to get drivers to increase their hazard perception (Horswill, 2016), found ways to get health personnel to increase their rate of hand washing (Grant & Hofmann, 2011), found ways to reduce health costs (Deangelis, 2010), found ways to decrease the occurrence of wrong-side surgery (McKinley et al., 2015; Zuger, 2015), have found out how to increase voter turnout (Bryan et al., 2011), and have found the answer to the age-old question of why children hate school (Willingham, 2010).

These applications of psychology have become so predictable and numerous that governments have formed special units to facilitate the use of behavioral science to foster broad public goals (Appelbaum, 2015; Author, 2016; Lewis, 2017). The United States established the Social and Behavioral Sciences Team (SBST) in 2014, and there is a parallel unit in the United Kingdom, the Behavioural Insights Team (BIT). These units have launched numerous projects based on behavioral science. For example, the SBST has projects aimed at student-loan default prevention, saving more for retirement, and inhibiting federal employees from texting while driving. The BIT has projects focused on facilitating taxpayer compliance and on timely vehicle registration behavior.

The "College Sophomore" Problem

The concerns of many people who question the "representativeness" of psychological findings focus on the subjects of the research rather than on the intricacies of the experimental design. We are confronting here what is sometimes called the *college sophomore problem*; that is, the worry that, because college sophomores are the subjects in an extremely large number of psychological investigations, the generality of the results is in question. Psychologists are concerned about the college sophomore issue because it is a real problem in certain areas of research. Nevertheless, it is important to

consider the problem in perspective and to understand that psychologists have several legitimate responses to this criticism. Here are three responses:

1. The college sophomore criticism does not *invalidate* past results, but simply calls for *more* findings that will allow assessment of the theory's generality. Adjustments in theory necessitated by contrary data from other groups can be made accurately only because we have the college sophomore data. The worst case, a failure to replicate, will mean that theories developed on the basis of college sophomore data are not necessarily wrong but merely incomplete.
2. In many areas of psychology, the college sophomore issue is simply not a problem because the processes investigated are so basic (the visual system, for example) that virtually no one would worry that their fundamental organization depends on the demographics of the subject sample. The functional organization of the brain and the nature of the visual systems of people in Montana tend to be very similar to those of people in Florida (or Argentina, for that matter).
3. Replication of findings ensures a large degree of geographic generality and, to a lesser extent, generality across socioeconomic factors, family variables, and early educational experience. As opposed to studies conducted 75 years ago, when the sample of university subjects participating would have come from an extremely elite group, research now goes on in universities that serve populations from a great variety of backgrounds.

It would be remiss, however, not to admit that the college sophomore issue is a real problem in certain areas of research in psychology. Nevertheless, psychologists are now making greater efforts to correct the problem. For example, developmental psychologists are almost inherently concerned about this issue. Each year hundreds of researchers in this area test dozens of findings and theories that were developed from studies of college subjects by performing the same research on subjects of different ages. The results from subject groups of different ages do not always replicate those from college students. Developmental psychology would be starkly boring if they did. But this sizable group of psychologists is busy building an age component into psychological theories, demonstrating the importance of this factor, and ensuring that the discipline will not end up with a large theoretical superstructure founded on a thin database derived from college students.

Psychologists also conduct cross-cultural research in order to assess the generality of the processes uncovered by researchers working only with North American subgroups. There are many instances in which cross-cultural comparisons have shown similar trends across cultures (e.g., Demetriou et al., 2005), but there are others in which cross-cultural research does not replicate the trends displayed by American college sophomores (e.g., Buchtel & Norenzayan, 2009; Henrich et al., 2010). However, when these discrepancies occur, they provide important information about the contextual dependence of theories and outcomes (Buchtel & Norenzayan, 2009; Henrich et al., 2010).

As previously mentioned, findings in cognitive psychology have met the basic test

As previously mentioned, findings in cognitive psychology have met the basic test of replicability. Many of the fundamental laws of information processing have been observed in dozens of laboratories all over the world. It is often not realized that if a psychologist at the University of Michigan obtains a finding of true importance, similar experiments will almost immediately be attempted at Stanford, Minnesota, Ohio State, Cambridge, Yale, Toronto, and elsewhere. Through this testing, we will soon know whether the finding is due to the peculiarities of the Michigan subjects or the study's experimental setting.

Cognitive, social, and clinical psychologists have also studied various human decision making strategies. Most of the original studies in this research area were done in laboratories, used college students as subjects, and employed extremely

artificial tasks. However, the principles of decision making behavior derived from these studies have been observed in a variety of nonlaboratory situations, including the prediction of closing stock prices by bankers, actual casino betting, prediction of patient behavior by psychiatrists, economic markets, military intelligence analysis, betting on NFL football games, estimation of repair time by engineers, estimation of house prices by realtors, business decision making, and diagnoses by physicians—and these principles are also now being applied in the very practical domain of personal finance (Kahneman, 2011; Lewis, 2017; Thaler, 2015; Zwieg, 2008).

The internet also provides a way for psychology to deal with the college sophomore problem (Germine et al., 2012; Maniaci & Rogge, 2014). Birnbaum (1999, 2004) ran a series of decision making experiments in the laboratory and by recruiting participants over the internet. The laboratory findings all replicated on the internet sample even though the latter was vastly more diverse—including 1,224 participants from 44 different countries. Gosling et al. (2004) studied a large internet sample of participants (361,703 people) and compared their performance with that of traditional samples in published studies. They found that the internet sample was more diverse with respect to gender, socioeconomic status, geographic region, and age. Importantly, they found that findings in many areas of psychology, such as personality theory, were similar on the internet when compared to traditional methods.

The Amazon Mechanical Turk (called MTurk) has been used extensively in recent psychological research to test subject samples that are at least somewhat different from college sophomores (DeSoto, 2016; Paolacci & Chandler, 2014; Stewart et al., 2015). The MTurk is an online marketplace of workers who are willing to complete experimental tasks for modest pay. The MTurk workers are considerably older than the college students used in most research (averaging over 30 years old), but they are atypical in other ways (they are less religious, underemployed, etc.). Nonetheless, many experimental effects that have been found in the lab are being tested on MTurk samples and are replicating with moderate frequency. Other internet sites such as Facebook are being used increasingly for psychological research (Kholodkov, 2013; Kosinski et al., 2015). These other sites also offer types of subjects very different from the typical college sophomore.

Of course, not all psychological findings replicate. On the contrary, replication failures do happen (Gilbert et al., 2016; Open Science Collaboration, 2015). The rate of replication failure in psychology has been an issue of intense discussion and debate during the last few years (Maxwell et al., 2015), as is the question of whether the rate in psychology is higher than that of other disciplines. This is a difficult question to answer, but it seems like failures of replication are less likely to be reported in psychology than in the physical sciences (Fanelli, 2010), indicating that psychology still has a way to go in upgrading its standards. Nevertheless, failures to replicate seem as prevalent in biology and medicine as they are in psychology (Author, 2013). The number of meta-analyses (discussed in Chapter 8) in psychology is increasing, a sign that the field is concerned with the consistency of its findings (Schmidt & Oh, 2016).

Nevertheless, there is encouraging evidence that a substantial number of psychological findings that originated in the laboratory do replicate in real-life field set-

psychological findings that originated in the laboratory do replicate in real-life field settings (although not all do). In the most comprehensive analysis to date, Mitchell (2012) meta-analyzed (see Chapter 8) data on 217 lab-field comparisons from various areas of psychology, such as industrial-organizational psychology, social psychology, and developmental psychology. He observed a substantial degree of correspondence between findings observed in the lab and those found in the field, but there was a large variation between different areas of psychology. Industrial-organizational psychology showed the highest degree of correspondence between lab and field, but social psychology was much lower. In 187 of 217 of the comparisons, the lab results and field results were in the same direction, but in 30 cases out of 217 the results in the lab were the opposite of those in the field. Among 30 reversals, the majority came from social psychology.

But how can any psychological findings be applied if replication failures sometimes occur? How can applications be justified if knowledge and theories are not established with certainty, when there is not complete agreement among scientists on all the details? This particular worry about the application of psychological findings is common because people do not realize that findings and theories in other sciences are regularly applied before they are firmly established. Of course, Chapter 2 should have made it clear that all scientific theories are subject to revision. If we must have absolutely certain knowledge before we can apply the results of scientific investigations, then no applications would ever take place. Applied scientists in all fields do their best to use the most accurate information available, realizing at the same time that the information is fallible.

Many nonscientists view medicine as much more scientific than psychology. But medicine has taken as long as psychology has to move from clinical impression to science-based practice (Lewis, 2017; Novella, 2015). Also, the uncertainty in the practice of medicine is no less than that in the practice of psychology. For example, key treatment-related findings in medicine often fail to replicate, diagnosis is often more a function of the physician than the disease, and new technologies often result in overtreatment that does not increase cure rates (Welch et al., 2012). Medical researchers still debate the benefits and harms of mammography at various ages (Kolata, 2014). The benefits and costs of a daily baby aspirin to prevent cardiovascular disease are still contested (Cuzick, 2015; Marks, 2015). Knowledge in psychology is probabilistic and uncertain—but the same is true in most other biosocial sciences.

The Real-Life and College Sophomore Problems in Perspective

Several issues have been raised in this chapter, and it is important to be clear about what has, and what has not, been said. We have illustrated that the frequent complaint about the artificiality of psychological research arises from a basic misunderstanding not only of psychology but also of basic principles that govern all sciences. Artificial conditions are not a drawback of experimental research. They are *deliberately* created so that we can pry variables apart.

We have also seen why people are concerned that psychologists do not use random samples in all their research and also why this worry is often unfounded. Finally, we have seen that a legitimate concern, the college sophomore problem, is sometimes overstated, particularly by those who are unfamiliar with the full range of activities and the diverse types of research that go on in psychology. The college sophomore problem has been an issue of great concern within psychology, and no psychologist is unaware of it. So, although we should not ignore the issue, we must also keep it in perspective.

Nevertheless, psychologists should always be concerned that their experimental conclusions not rely too heavily on any one method or particular subject population. The next chapter deals with this very point. Indeed, some areas of psychology *are*

plagued by a college sophomore problem (Jaffe, 2005; Henrich et al., 2010). Cross-cultural psychology, an antidote to the college sophomore problem, is not yet fully integrated with psychology as a whole (Wang, 2017). As mentioned previously, the field's replication failure rate, while typical of a behavioral science, is still worrisome. Some psychologists have pointed out that there is too much flexibility in our data analysis procedures (Simmons et al., 2011). Finally, as will be discussed in Chapter 12, there is a growing problem in psychology that too many of its researchers (particularly in universities) share preexisting biases, especially political ones (Duarte et al., 2015; Inbar & Lammers, 2012; Jussim et al., 2016; Lukianoff & Haidt, 2015; Tetlock, 2012).

Summary

Some psychological research is applied work in which the goal is to relate the results of the study directly to a particular situation. In such applied research, in which the results are intended to be extrapolated directly to a naturalistic situation, questions of the randomness of the sample and the representativeness of the conditions are important because the findings of the study are going to be applied directly. However, most psychological research is not of this type. It is basic research designed to test theories of the underlying mechanisms that influence behavior. In most basic research, the findings are applied only indirectly through modifications in a theory that will at some later point be applied to some practical problem. In basic research of this type, random sampling of subjects and representative situations are not an issue because the emphasis is on testing the universal prediction of a theory. In fact, artificial situations are deliberately constructed in theory-testing basic research because (as described in the previous chapter) they help to isolate the critical variable for study and to control extraneous variables. Thus, the fact that psychology experiments are “not like real life” is a strength rather than a weakness.

Chapter 8

Avoiding the Einstein Syndrome: The Importance of Converging Evidence

Learning Objectives

8.1 Compare the breakthrough and gradual-synthesis models of scientific progress

8.2 Describe the principle of converging evidence in evaluating experiments and testing theories

8.3 Explain how multiple research methods are used to arrive at a scientific consensus

8.4 Explain why meta-analyses is used to draw conclusions in psychology

"Biological Experiment Reveals the Key to Life," "New Breakthrough in Mind Control," "California Scientist Discovers How to Postpone Death"—as you can see, it is not difficult to parody the "breakthrough" headlines of the media (including print media, television, and the internet). Because such headlines regularly come from the most irresponsible quarters of the media, it should not be surprising that most scientists recommend that they be approached with skepticism. The purpose of this chapter, though, is not only to warn against the spread of misinformation via exaggeration or to caution that the source must be considered when evaluating reports of scientific advances. In this chapter, we also want to develop a more complex view of the scientific process than was presented in earlier chapters. We shall do this by elaborating on the ideas of systematic empiricism and public knowledge that were introduced in Chapter 1.

The breakthrough headlines in the media obscure an understanding of psychology

and other sciences in many ways. One particular misunderstanding that arises from breakthrough headlines is the implication that all problems in science are solved when a single, crucial experiment completely decides the issue, or that theoretical advance is the result of a single critical insight that overturns all previous knowledge. Such a view of scientific progress fits in nicely with the operation of the news media and the internet, in which history is tracked by presenting separate, disconnected events in bite-sized units. It is also a convenient format for the Hollywood entertainment industry, where events must have beginnings and satisfying endings that resolve ambiguity. However, this is a gross caricature of scientific progress and, if taken too seriously, leads to misconceptions about scientific advancement and impairs the ability to evaluate the extent

of scientific knowledge on a given issue. In this chapter, we will discuss two principles of science—the connectivity principle and the principle of converging evidence—that describe scientific progress much more accurately than the breakthrough model.

The Connectivity Principle

In denying the validity of the "great-leap" or crucial-experiment model of all scientific progress, we do not wish to argue that such critical experiments and theoretical advances never occur. On the contrary, some of the most famous examples in the history of science represent just such occurrences. The development of the theory of relativity by Albert Einstein is by far the most well known. Here, a reconceptualization of such fundamental concepts as space, time, and matter was achieved by a series of remarkable theoretical insights.

However, the monumental nature of Einstein's achievement has made it the dominant model of scientific progress in the public's mind. This dominance is perpetuated because it fits in nicely with the implicit "script" that the media use to report most news events. More nonsense has been written about relativity theory than perhaps any other idea in all of history (no, Einstein did *not* prove that "everything is relative"). Of course, our purpose is not to deal with all of these fallacies here. There is one, however, that will throw light on our later discussions of theory evaluation in psychology.

The reconceptualization of ideas about the physical universe contained in Einstein's theories is so fundamental that popular writing often treats it as if it were similar to conceptual changes in the arts (a minor poet is reevaluated and emerges with the status of a genius; an artistic school is declared dead). Such presentations ignore a basic difference between conceptual change in the arts and in the sciences.

Conceptual change in science obeys a principle of *connectivity* that is absent or, at least, severely limited in the arts (Bronowski, 1977; Haack, 2007). That is, a new theory in science must make contact with previously established empirical facts. To be considered an advance, it must not only explain new facts but also account for old ones. The theory may explain old facts in a way quite different from that of a previous theory, but explain them it must. This requirement ensures the cumulative progress of science. Genuine progress does not occur unless the realm of our explanatory power has been widened. If a new theory accounts for some new facts but fails to account for a host of old ones, it will not be considered an advance over the old theories and, thus, will not immediately replace them.

Despite the startling reconceptualizations in Einstein's theories (clocks in motion running slower, mass increasing with velocity, etc.), they did maintain the principle of connectivity. In rendering Newtonian mechanics obsolete, Einstein's theories did not negate or render meaningless the facts about motion on which Newton's ideas were based. On the contrary, at low velocities the two theories make essentially the same predictions. Einstein's conceptualization is superior because it accounts for a wide variety of new, sometimes surprising, phenomena that Newtonian mechanics cannot accommodate. Thus, even Einstein's theories, some of the most startlingly new and fundamental

date. Thus, even Einstein's theories, some of the most startlingly new and fundamental reconceptualizations in the history of science, maintain the principle of connectivity.

A Consumer's Rule: Beware of Violations of Connectivity

The breakthrough model of scientific progress—what we might call the Einstein syndrome—leads us astray by implying that new discoveries violate the principle of connectivity. This implication is dangerous because, when the principle of connectivity is abandoned, the main beneficiaries are the purveyors of pseudoscience and bogus theories. Such theories derive part of their appeal and much of their publicity from the fact that they are said to be startlingly new. "After all, wasn't relativity new in its day?" is usually

the tactic used to justify novelty as a virtue. Of course, the data previously accumulated in the field that the pseudoscientists wish to enter would seem to be a major obstacle. Actually, however, it presents only a minor inconvenience because two powerful strategies are available to dispose of it. One strategy that we have already discussed (see Chapter 2) is to explain the previous data by making the theory unfalsifiable and, hence, useless.

The second strategy is to dismiss previous data by declaring them irrelevant. This dismissal is usually accomplished by emphasizing what a radical departure the new theory represents. The phrases "new conception of reality" and "radical new departure" are frequently used. Vague references to quantum theory are often thrown in to suggest that the new theory is deep and profound (DeBakcsy, 2014; Hassani, 2016). The real sleight of hand, though, occurs in the next step of the process. The new theory is deemed so radical that experimental evidence derived from the testing of other theories is declared irrelevant. Only data that can be conceptualized within the framework of the new theory are to be considered; that is, the principle of connectivity is explicitly broken. Obviously, because the theory is so new, such data are said to not yet exist. And there you have it: a rich environment for the growth of pseudoscience. The old, "irrelevant" data are gone, and the new, relevant data do not exist. The scam is easily perpetrated because the Einstein syndrome obscures the principle of connectivity, the importance of which is ironically illustrated by Einstein's theories themselves.

It is likewise with psychology. A new theory that denied the existence of classical and operant conditioning would never develop in psychology because it would not connect with what else is known in behavioral science. Recall the discussion of facilitated communication in Chapter 6. It breaks the principle of connectivity because it would require that we overturn basic knowledge in fields as diverse as neurology, genetics, and cognitive psychology. This hypothesized cure shows no connectivity with the rest of science.

Consider another example from psychology. Imagine two specific treatments have been developed to remediate the problems of children with extreme reading difficulties. No direct empirical tests of efficacy have been carried out using either treatment. The first, Treatment A, is a training program to facilitate the awareness of the segmental nature of language at the phonological level. The second, Treatment B, involves giving children training in vestibular sensitivity by having them walk on balance beams while blindfolded. Even if there was no prior evidence on either treatment, one of them has the edge when it comes to the principle of connectivity. Treatment A makes contact with a broad consensus in the research literature that children with reading difficulties are hampered because of insufficiently developed awareness of the segmental structure of language (Hulme & Snowling, 2013; Seidenberg, 2017). Treatment B is not connected to any corresponding research literature consensus. This difference in connectivity dictates that Treatment A is a better choice.

Neurologist Steven Novella (2015) makes the same point about complementary and alternative medicine. Saying that complementary and alternative medicine lacks empirical evidence—which it does (Dorlo et al., 2015; Mielczarek & Engler, 2013; Swan et al., 2015)—is in one sense too generous. Novella (2015) points at the fact that most of

these remedies do not *deserve* to have experimental tests conducted on them—because they display no connectivity with the rest of science.

The "Great-Leap" Model Versus the Gradual-Synthesis Model

The tendency to view the Einsteinian revolution as typical of what science is tempts us to think that all scientific advances occur in giant leaps. The problem is that people tend to generalize such examples into a view of the way all scientific progress *should* take place. In fact, many areas in science have advanced not by single, sudden breakthroughs but by series of fits and starts that are less easy to characterize.

There is a degree of fuzziness in the scientific endeavor that most of the public is unaware of. Experiments rarely completely decide a given issue, supporting one theory and ruling out all others. New theories are rarely clearly superior to all previously existing competing conceptualizations. Issues are most often decided not by a critical experiment, as movies about science imply, but when the community of scientists gradually begins to agree that the preponderance of evidence supports one theory rather than another. The evidence that scientists evaluate is not the data from a single experiment that has finally been designed in the perfect way. Instead, scientists most often must evaluate data from literally dozens of experiments, each containing some flaws but each providing a small part of the answer. This alternative model of scientific progress has been obscured because the Einstein syndrome creates in the public a tendency to think of all science by reference to physics, to which the great-leap model of scientific progress is perhaps most applicable.

Consider the rapid advances in genetics and molecular biology that have occurred in the last hundred years. These advances have occurred not because one giant, Einstein, came onto the scene at the key moment to set everything straight. Instead, dozens of different insights based on hundreds of experiments have contributed to the modern synthesis in biology. These advances occurred not by the instantaneous recognition of a major conceptual innovation, but by long, drawn-out haggling over alternative explanations, each of which had partial support. It took over ten years of inconclusive experimentation, along with much theoretical speculation, argument, and criticism, for scientists to change their view about whether genes were made of protein or nucleic acid. The consensus of opinion changed, but not in one great leap.

Science is a cumulative endeavor that respects the principle of connectivity. It is characterized by the participation of many individuals, whose contributions are judged by the extent to which they further our understanding of nature. No single individual can dominate discourse simply by virtue of his or her status. Science rejects claims of "special knowledge" available to only a few select individuals. This rejection, of course, follows from our discussion of the public nature of science in Chapter 1. By contrast, pseudosciences often claim that certain authorities or investigators have a "special" access to the truth.

We have presented two ideas here that provide a useful context for understanding the discipline of psychology. First, no experiment in science is perfectly designed. There is a degree of ambiguity in the interpretation of the data from any one experiment. Scientists often evaluate theories not by waiting for the ideal or crucial experiment to appear, but by assessing the overall trends in a large number of experiments—each with different limitations. Second, many sciences have progressed even though they are without an Einstein. Their progress has occurred by fits and starts, rather than by discrete stages of grand Einsteinian syntheses. Also like psychology, many other sciences are characterized instead by growing mosaics of knowledge that lack a single integrating theme.

Converging Evidence: Progress Despite Flaws

The previous discussion has led to a principle of evidence evaluation of much importance in psychology. This idea is sometimes called the *principle of converging evidence* (or *converging operations*). Scientists and those who apply scientific knowledge must often make a judgment about where the preponderance of evidence points. When this is the case, the principle of converging evidence is an important tool. We will explore two ways of expressing the principle, one in terms of experiments with limitations and the other in terms of theory testing.

There are always a number of ways in which an experiment can go wrong (or become *confounded*, to use the technical term). However, a scientist with much experience in working on a particular problem usually has a good idea of what the most likely confounding factors are. Thus, when surveying the research evidence, scientists are usually aware of the critical flaws in each experiment. The idea of converging evidence, then, tells us to examine the pattern of flaws running through the research literature because the nature of this pattern can either support or undermine the conclusions that we wish to draw.

For example, suppose the findings from a number of different experiments were largely consistent in supporting a particular conclusion. Given the imperfect nature of experiments, we would go on to evaluate the extent and nature of the limitations in these studies. If all the experiments were limited in a similar way, this circumstance would undermine confidence in the conclusions drawn from them because the consistency of the outcome may simply have resulted from a particular flaw that all the experiments shared. On the other hand, if all the experiments were limited in *different* ways, our confidence in the conclusions would be increased because it is less likely that the consistency in the results was due to a contaminating factor that confounded all the experiments.

Each experiment helps to correct errors in the design of other experiments, and when evidence from a wide *range* of experiments points in a similar direction, then the evidence has converged. A reasonably strong conclusion is justified even though no one experiment was perfectly designed. Thus, the principle of converging evidence urges us to base conclusions on data that arise from a number of slightly different experimental sources. The principle allows us to draw stronger conclusions because consistency that has been demonstrated in such a context is less likely to have arisen from the peculiarities of a single type of experimental procedure.

The principle of converging evidence can also be stated in terms of theory testing. Research is highly convergent when a series of experiments consistently supports a given theory while collectively eliminating the most important competing theory. Although no single experiment can rule out all alternative explanations, taken collectively a series of partially diagnostic experiments can lead, if the data patterns line up in a certain way, to a strong conclusion.

For example, suppose that five different theoretical accounts (call them A, B, C, D, and E) of a given set of phenomena exist at one time and are investigated in a series of experiments. Suppose that one experiment represents a strong test of theories A, B, and C, and that the data largely refute theories A and B and support C. Imagine also that another experiment is a particularly strong test of theories C, D, and E, and that the data largely refute theories D and E and support C. In such a situation, we would have strong converging evidence for theory C. Not only do we have data supportive of theory C, but we have data that contradict its major competitors. Note that no one experiment tests all the theories, but taken together, the entire set of experiments allows a strong inference. The situation might be depicted like the following:

	Theory A	Theory B	Theory C	Theory D	Theory E
Experiment 1	refuted	refuted	supported	untested	untested
Experiment 2	untested	untested	supported	refuted	refuted

	Theory A	Theory B	Theory C	Theory D	Theory E
Conclusion	refuted	refuted	supported	refuted	refuted

By contrast, if both experiments represented strong tests of B, C, and E, and the data of both experiments strongly supported C and refuted B and E, the overall support for theory C would be less strong than in our previous example. The reason

is that, although data supporting theory C have been generated, there is no strong evidence ruling out two viable alternative theories (A and D). The situation would be something like the following:

	Theory A	Theory B	Theory C	Theory D	Theory E
Experiment 1	untested	refuted	supported	untested	refuted
Experiment 2	untested	refuted	supported	untested	refuted

	Theory A	Theory B	Theory C	Theory D	Theory E
Conclusion	untested	refuted	supported	untested	refuted

Thus, research is highly convergent when a series of experiments consistently supports a given theory while collectively eliminating the most important competing explanations. Although no single experiment can rule out all alternative explanations, taken collectively a series of partially diagnostic experiments can lead to a strong conclusion if the data converge in the manner of our first example.

Finally, the introduction of the idea of converging evidence allows us to dispel a misconception that may have been fostered by our oversimplified discussion of falsifiability in Chapter 2. That discussion may have seemed to imply that a theory is falsified when the first piece of evidence that disconfirms it comes along. This is not the case, however. Just as theories are confirmed by converging evidence, they are also disconfirmed by converging results.

Knowledge of the principle of converging evidence is what leads retired physician Harriet Hall (2013) to warn us to be skeptical of the phrase “new study shows” when we see it in the media or on the internet. You know the type of thing: New Study Shows that People Who Eat Kumquats Live 40 percent Longer. The reason to be skeptical should be clear to you by now: no *single* study *shows* anything! Before we derive a conclusion, many studies must be amalgamated together and we must assess whether they converge. Renowned cognitive psychologist Steven Pinker echoes this point: “There’s a habit among science journalists to treat a single experiment as something that is newsworthy. But a single study proves very little. Readers have been led to expect shocking discoveries from a discipline that depends on slow, stutter-step progress” (p. 92, Kachka, 2012). Joe Palca, a science reporter for National Public Radio (who has a PhD in psychology), agrees with Pinker when he says that “science goes in increments, and the media goes in leaps and bounds” (p. 19, Miller, 2016).

Types of Converging Evidence

The reason for stressing the importance of convergence is that conclusions in psychology are often based on the principle of converging evidence. There is certainly nothing unique or unusual about this fact (conclusions in many other sciences rest not on single, definitive experimental proofs, but on the confluence of dozens of fuzzy

experiments). But there are reasons that this might be especially true of psychology. Experiments in psychology are usually of fairly low diagnosticity. That is, the data that support a given theory usually rule out only a small set of alternative explanations, leaving many additional theories as viable candidates. As a result, strong conclusions are usually possible only after data from a very large number of studies have been collected and compared.

Better public understanding will come about if psychologists openly acknowledge this fact and then take pains to explain just what follows from it. Psychologists should admit that, although a science of psychology exists and is progressing, progress is slow, and our conclusions come only after a sometimes excruciatingly long period of research amalgamation and debate.

Media claims (whether in print, television, or on the internet) of breakthroughs should always engender skepticism, but this is especially true of psychological claims. For example, it sometimes seems like the media announces a new cure for autism about every three months. But these claims have been continuously occurring for over 20 years now. Why are we still announcing a cure for autism when one was announced 20 years ago?…and 19 years ago?…and 18 years ago?…etc. This of course suggests that the announcement 20 years ago was not a true cure at all. Perhaps it was a bogus claim. More likely, it was just a small step in the long road of scientific progress that will lead to a convergence of evidence regarding this condition. But these premature media reports wrongly imply that research on autism is noncumulative—that researchers are not slowly building knowledge, but instead are searching for a magic bullet.

It is likewise with a research specialty area that I worked in during the early part of my career—the psychology of reading and reading disability. As with autism, a "cure" (magic bullet) for dyslexia has been announced in the media nearly yearly since about 1990! For some examples, I haphazardly perused a bulging clip file of articles I have collected on these premature announcements and came up with the November 22, 1999, issue of *Newsweek* magazine featuring on its cover a lead article titled: Dyslexia: New Hope for Kids Who Can't Read (Kantrowitz & Underwood, 1999). And here is an article in the February 26, 2001, *National Post* (Canada) newspaper with an article titled: Unravelling Dyslexia's Riddle (Gains, 2001). And here comes the cover of the July 28, 2003, *Time* magazine with a cover titled Overcoming Dyslexia—What New Brain Science Reveals (Gorman et al., 2003). And finally, fairly close to the present, here comes *Newsweek* again, on March 31, 2016, with an article titled Electric Shocks Help Dyslexic Children Read Faster (Cuthbertson, 2016). I'll stop there. There was no magic bullet in any of these articles. My point is not that the research reported in these articles was bad or wrong. The important point to understand though is that the media sources exaggerated the "magic bullet" nature of the studies reported. They were not "cures," but instead were part of the slow progress that is definitely being made in the area of reading disability (Seidenberg, 2017).

The media does the same thing with attention deficit hyperactivity disorder (ADHD)—announces startling new discoveries (magic bullets) vastly prematurely. This tendency to prematurely report breakthroughs in the media has been studied in the ADHD area. A group of researchers studied the ten most publicized scientific articles on ADHD over a ten year period in the 1990s (Gonon et al., 2012). These ten articles resulted in 347 newspaper reports (typical title: "Hyperactivity Linked to Genetic Defect"). The researchers then looked at the next decade of research to see if the ten findings replicated. What they found confirms our fears about premature reports in the media. Only two of the ten were strongly replicated. Six failed to replicate completely. Two others showed attenuated findings (findings not as strong as in the original report). In short, these studies did not deserve to be publicized as "breakthroughs" or magic bullets. They were instead just small, confusing (and sometimes wrong) steps toward an eventual understanding of ADHD. Indeed, the premature

hyping of studies such as these has, ironically, been termed "Journalistic Deficit Disorder" (Reporting Science, 2012).

In psychology we have to walk a very fine line. For example, we must resist the temptation to regard a particular psychological hypothesis as "proven" when the evidence surrounding it is still ambiguous. This skeptical attitude has been reinforced in several chapters of this book. The cautions against inferring causation from correlation and against accepting testimonial evidence have served as examples. At the same time, we should not overreact to the incompleteness of knowledge and the tentativeness of conclusions by doubting whether firm conclusions in psychology will ever be reached. Nor should we be tempted by the irrational claim that psychology cannot be a science. From this standpoint, the principle of converging evidence can be viewed as a counterweight to the warnings against overinterpreting tentative knowledge.

Convergence allows us to reach many reasonably strong conclusions despite the flaws in all psychological research.

The best way to see the power of the principle of converging evidence is to examine some areas in psychology where conclusions have been reached by the convergence of evidence. Let's consider an example. A research problem that illustrates the importance of the principle of converging evidence is the question of whether exposure to violent television programming increases children's tendencies toward aggressive behavior. There is now a scientific consensus on this issue: The viewing of violent programming (on television, in movies, or in streaming video) does appear to increase the probability that children will engage in aggressive behavior. The effect is not extremely large, but it is real. Again, the confidence that scientists have in this conclusion derives not from a single definitive study, but from the convergence of the results of dozens of different investigations (Bushman et al., 2016; Carnagey et al., 2007; Feshbach & Tangney, 2008; Fischer et al., 2011b). This research conclusion holds for violent video games as well as television and movies (Calvert et al., 2017; Carnagey et al., 2007), but the effect there seems to be fairly small as well (Ferguson, 2013; Furuya-Kanamori & Doi, 2016). The general research designs, subject populations, and specific techniques used in these investigations differed widely, and as should now be clear, these differences are a strength of the research in this area, not a weakness.

Television network executives and video game industry executives, naturally resistant to evidence of the negative effects of their industries on children, have carried on a campaign of misinformation that capitalizes on the public's failure to realize that research conclusions are based on the convergence of many studies rather than on a single critical demonstration that decides the issue (Seethaler, 2009). The television networks and video game makers continually single out individual studies for criticism and imply that the general conclusion is undermined by the fact that each study has demonstrated flaws. But it is not commonly recognized that researchers often candidly *admit* the flaws in a given study. The critical difference is that researchers reject the implication that admitting a flaw in a given study undermines the general scientific consensus on the effects of televised violence on aggressive behavior. The reason is that the general conclusion derives from a convergence. Research without the specific flaws of the study in question has produced results pointing in the same direction. This research may itself have problems, but other studies have corrected for these and have also produced similar results.

For example, very early in the investigation of this issue, evidence of the correlation between the amount of violent programming viewed and aggressive behavior in children was uncovered. It was correctly pointed out that this correlational evidence did not justify a causal conclusion. Perhaps a third variable was responsible for the association, or perhaps more aggressive children chose to watch more violent programming (the directionality problem).

But the conclusion of the scientific community is not based on this correlational evidence alone. There are more complex correlational techniques than the simple mea-

surement of the association between two variables, and these correlational techniques allow some tentative conclusions about causality (one, that of partial correlation, was mentioned in Chapter 5). One of these techniques involves the use of a longitudinal design in which measurements of the same two variables—here, television violence and aggression—are taken at two different times. Certain correlational patterns suggest causal connections. Studies of this type have been conducted, and the pattern of results suggested that viewing violent programming did tend to increase the probability of engaging in aggressive behavior later in life.

Again, it is not unreasonable to counter that these longitudinal correlational techniques are controversial, because they are. The important point is that the conclusion of a causal connection between televised violence and aggressive behavior does not depend entirely on correlational evidence, either simple or complex, because

numerous laboratory studies have been conducted in which the amount of televised violence was manipulated rather than merely assessed. In Chapter 6, we discussed how the manipulation of a variable, used in conjunction with other experimental controls such as random assignment, prevents the interpretation problems that surround most correlational studies. If two groups of children, experimentally equated on all other variables, show different levels of aggressive behavior, and if the only difference between the two is that one group viewed violent programming and one did not, then we are correct in inferring that the manipulated variable (televised violence—the independent variable) caused the changes in the outcome variable (aggressive behavior—the dependent variable). This result has occurred in the majority of studies.

These studies have prompted some to raise the "it's-not-real-life" argument discussed in the previous chapter and to use the argument in the fallacious way discussed in that chapter. In any case, the results on the effects of television violence are not peculiar to a certain group of children because these results have been replicated in different regions of the United States and in several countries around the world. The specific laboratory setup and the specific programs used as stimuli have varied from investigation to investigation, yet the results have held up.

Importantly, the same conclusions have been drawn from studies conducted in the field rather than in the laboratory. A design discussed in Chapter 6, known as the *field experiment* has been used to investigate the televised-violence/aggressive-behavior issue. The existence of this type of design reminds us to avoid assuming a necessary link between experimental design and experimental setting. People sometimes think that studies that manipulate variables are conducted only in laboratories and that correlational studies are conducted only in the field. This assumption is incorrect. Correlational studies are often conducted in laboratories, and variables are often manipulated in nonlaboratory settings. Although they sometimes require considerable ingenuity to design, field experiments (several of which were mentioned in Chapter 6), in which variables are manipulated in nonlaboratory settings, are becoming more common in psychology.

Of course, field experiments themselves have weaknesses, but many of these weaknesses are the strengths of other types of investigation. In summary, the evidence linking the viewing of televised violence to increased probabilities of aggressive behavior in children does not rest only on the outcome of one particular study or even on one generic type of study.

The situation is analogous to the relationship between smoking and lung cancer. Smokers are 15 times more likely to die from lung cancer than nonsmokers (Gigerenzer et al., 2007). In the past, cigarette company executives often attempted to mislead the public by implying that the conclusion that smoking causes lung cancer rested on some specific study, which they would then go on to criticize (Offit, 2008). Instead, the conclusion is strongly supported by a wealth of converging evidence. The convergence of data from several different types of research is quite strong and will not be changed substantially by the criticism of one study.

Actually it is appropriate to discuss here a medical problem like the causes of lung cancer. Most issues in medical diagnosis and treatment are decided by an amalgamation of converging evidence from many different types of investigations. For example, medical science is confident of a conclusion when the results of epidemiological studies (field studies of humans in which disease incidence is correlated with many environmental and demographic factors), highly controlled laboratory studies using animals, and clinical trials with human patients all converge. When the results of all these types of investigation point to a similar conclusion, medical science feels assured of the conclusion, and physicians feel confident in basing their treatment on the evidence.

However, each of the three different types of investigation has its drawbacks. Epidemiological studies are always correlational, and the possibility of spurious links

between variables is high. Laboratory studies can be highly controlled, but the subjects are often animals rather than humans. Clinical trials in a hospital setting use human subjects in a real treatment context, but there are many problems of control because of placebo effects and the expectations of the medical treatment team that deals with the patients. Despite the problems in each type of investigation, medical researchers are justified in drawing strong conclusions when the data from all the different methods converge strongly, as in the case of smoking and lung cancer. Just such a convergence also justifies the conclusions that psychologists draw from the study of a behavioral problem like the effect of televised violence on aggressive behavior.

Sometimes the principle of converging evidence is unknown to people. Other times it seems to be consciously ignored in order to advance a political agenda or an agenda of financial advancement. Certainly the cigarette company experts and senior executives who tried to confuse the public's understanding of the converging evidence that smoking caused lung cancer probably were aware of the convergence principle and wished to obscure it from the public.

An example similar to the smoking/lung cancer case is occurring right at the present time. There is a strong convergence in science indicating that talking on a cell phone while driving (as well as the distraction from electronic dashboard devices while driving) is extremely dangerous and an important cause of car crashes (even with hands-free phones). The convergence comes from lab studies, field studies, correlational studies, true experiments, and from connectivity with attentional theory in cognitive science. Yet cell phone companies and auto companies—like the cigarette companies before them—are attempting to obscure from the public the fact that the science surrounding this conclusion is highly convergent (Insurance Institute for Highway Safety, 2005; Kunar et al., 2008; Levy et al., 2006; McEvoy et al., 2005; Richtel, 2014; Strayer et al., 2016; Strayer & Drews, 2007). The technology companies and automobile companies ignore the science even more when they try to get a competitive edge by installing more interactive electronic features in cars. Apple's CarPlay technology and Google's Android Auto are particularly troublesome developments (Chaker, 2016; White, 2014), given the science on driver distraction. The technology companies and automobile companies continue to ignore the science of driver risk. These technology-caused deaths are preventable with electronic fixes now available, but these modern companies continue to act like the cigarette companies did years ago in their reluctance to deal with known consumer risks (Leonhardt, 2017).

Scientific Consensus

The problem of assessing the impact of televised violence is typical of how data finally accumulate to answer questions in psychology. Particularly in areas of pressing social concern, it is wise to remember that the answers to these problems emerge only slowly, after the amalgamation of the results from many different experiments. To put things

in the form of a simple rule, when evaluating empirical evidence in the field of psychology, think in terms of *scientific consensus* rather than breakthrough—in terms of *gradual synthesis* rather than great leap.

The failure to appreciate the "consensus rather than breakthrough" rule has impeded the public's understanding of the evidence that human activity is a contributor to global warming (Cook, 2016; Powell, 2015). In fact, there is not much scientific controversy about this conclusion (in its broadest sense), because the conclusion does not rest on a single study. There were over 900 global climate-change papers published between 1993 and 2003, and they overwhelmingly converged on the conclusion that human activity was involved in global warming (Oreskes & Conway, 2011). No single study was definitive in establishing the conclusion so, obviously, undermining a single study would not change the conclusion at all. Note, however, that establishing the conclusion in a broad

sense does not necessarily dictate what should be *done* in response to the conclusion. What is to be done is a political judgment. The fact itself is in the realm of science. The fact does not necessarily dictate a *particular* policy response—or any response at all.

Methods and the Convergence Principle

The convergence principle also implies that we should expect many different meth- [illegible]. [illegible]. Psychology has long been criticized for relying too heavily on laboratory-based experimental techniques. Nevertheless, an unmistakable trend in recent years has been to expand the variety of methods used in all areas of psychology. Researchers have turned to increasingly imaginative field designs in search of converging evidence to support their theories.

As an example, consider the voluminous research done on what has been called the [illegible], that is, [illegible] (Fischer et al., 2011a; Thomas et al., 2016). Probability of helping can sometimes go down as more potential helpers are present. The early investigators of this phenomenon were well aware that their conclusions would be tenuous if they were based only on the responses of individuals who witnessed emergencies after reporting to a laboratory to participate in an experiment. Therefore, in an early famous study of this effect, researchers found a cooperative liquor store that agreed to have fake robberies occur in the store 96 different times. While the cashier was in the back of the store getting some beer for a "customer," who was actually an accomplice of the experimenter, the "customer" walked out the front door with a case of beer. This was done in the view of either one or two real customers who were at the checkout counter. The cashier then came back and asked the customers, "Hey, what happened to that man who was in here? Did you see him leave?" thus giving the customers a chance to report the theft. Consistent with the laboratory results, the presence of another individual inhibited the tendency to report the theft.

Many of the principles of probabilistic decision making to be discussed in Chapter 10 originated in the laboratory but have also been tested in the field. For example, researchers have used laboratory-derived principles to explain the way that physicians, stockbrokers, jurors, economists, and gamblers reason probabilistically in their environments (Kahneman, 2011; Lewis, 2017; Thaler, 2015; Zweig, 2008). The convergence of laboratory and nonlaboratory results has also characterized several areas of educational psychology. For example, both laboratory studies and field studies of different curricula have indicated that early phonics instruction facilitates the acquisition of reading skill (Ehri et al., 2001; Seidenberg, 2017; Willingham, 2017).

It should be remembered that [illegible], [illegible]

[redacted]. [redacted]—that the hypothesis that was originally posited cannot be supported. [redacted]. It has long been thought there was a way for teachers to measure each child's “learning style.” Now I won't mention a specific style here, because [redacted]). In any case, teachers are then supposed to be able to “teach to” these styles—resulting in higher achievement for all. (It is sometimes also claimed that students will all achieve much more equally if this is done.) [redacted] (Hood, 2017; Kirschner & van Merrienboer, 2013; Pashler et al., 2009). [redacted].

The Progression to More Powerful Methods

[redacted]. For example, interest in a particular hypothesis may originally stem from a particular case study of unusual interest. As we discussed in Chapter 4, this is the proper role of case studies: to suggest hypotheses for further study with more powerful techniques and to motivate scientists to apply more rigorous methods to a research problem. Thus, following the case studies, researchers undertake correlational investigations to verify whether the link between variables is real rather than the result of the peculiarities of a few case studies. If the correlational studies support the relationship between relevant variables, researchers will attempt experiments in which variables are manipulated in order to isolate a causal relationship between the variables.

[redacted], then, [redacted]. What type of study is most appropriate often depends on how advanced the research problem is. Past-president of the Association for Psychological Science Doug Medin (2012) reminds us that “some well-established areas of research may be like Phase III clinical trials, in which the methods and measures are settled issues and the only concern is with assessing effect size. Other areas, however, may rely on open-ended tasks in which the dependent variable cannot and typically should not be specified in advance” (p. 6).

Discussing the idea of the progression through the more powerful research methods provides us with a chance to deal with a misconception that some readers may have derived from Chapter 5—that is, that correlational studies are not useful in science. [redacted]. [redacted] (West, 2009). First, [redacted]. Second, [redacted]. That is, [redacted]. Third, [redacted]. We discussed in Chapter 5 the complex correlational technique of partial correlation, in which it is possible to test whether a particular third variable is accounting for a relationship.

Perhaps most important, however, [redacted]. [redacted]. This circumstance, again, is not unique to psychology. Astronomers obviously cannot manipulate all the variables affecting the objects they study, yet they are able to arrive at conclusions.

An example of the evolution of research methods in health psychology is the work concerning the link between the type A behavior pattern and coronary heart disease (Chida & Hamer, 2008; Martin et al., 2011; Matthews, 2013). The original observations that led to the development of the concept of the type A behavior pattern occurred when two cardiologists thought they noticed a pattern in the behavior of some of their coronary patients that included a sense of time urgency, free-floating hostility, and extremely competitive striving for achievement. Thus, the idea of the type A personality originated in a few case studies made by some observant physicians. These case studies suggested the concept, but they were not taken as definitive proof of the hypothesis that a particular type of behavior pattern is a partial cause of coronary heart disease. Proving the idea required more than just the existence of a few case studies. It involved decades of work by teams of cardiologists and psychologists.

The research quickly moved from merely accumulating case studies, which could never establish the truth of the hypothesis, to more powerful methods of investigation. Researchers developed and tested operational definitions of the type A concept. Large-scale epidemiological studies established a correlation between the presence of type A behavior and the incidence of coronary heart disease. The correlational work then became more sophisticated. Researchers used complex correlational techniques to track down potential third variables. The relation between type A behavior and heart attacks could have been spurious because the behavior pattern was also correlated with one of the other traditional risk factors (such as smoking, obesity, or serum cholesterol level). However, results showed that type A behavior was a significant independent predictor of heart attacks. When other variables were statistically partialed out, there was still a link between the type A behavior pattern and coronary heart disease.

Finally, researchers undertook experimental studies with manipulated variables to establish whether a causal relationship could be demonstrated. Some of the studies attempted to test models of the physiological mechanisms that affected the relationship and used animals as subjects—what some might call "not real life." Another experimental study used human subjects who had had a heart attack. These subjects were randomly assigned to one of two groups. One group received counseling designed to help them avoid traditional risky behavior such as smoking and eating fatty foods. The other group received this counseling and were also given a program designed to help them reduce their type A behavior. Three years later, there had been significantly fewer recurrent heart attacks among the patients given the type A behavior counseling.

In short, the evidence converged to support the hypothesis of the type A behavior pattern as a significant causal factor in coronary heart disease. The investigation of this problem provides a good example of how [redacted].

Another lesson we can draw from this example is that [redacted]—an issue first raised in Chapter 3 when we discussed operational definitions. Recent research seems to indicate that it is oversimplifying to talk about the connection between heart attacks and the type A behavior pattern *as a whole.* The reason is that only certain components of the pattern (particularly antagonistic hostility) appear to be linked to coronary heart disease (Chida & Hamer, 2008; Matthews, 2013). Thus, we have an example of how science uncovers increasingly specific relationships as it progresses and how theoretical concepts become elaborated.

There is a final point to note in our discussion of scientific consensus. [redacted]. [redacted] (Vyse, 2017). Scientists are of course free to sign whatever petition they want about whatever social or political issue,

but such a document is not what we are talking about in this chapter when we talk about a scientific consensus. We will see in Chapter 12 that the American Psychological Association itself has been guilty of taking positions on social issues that are only loosely connected (or not connected at all) to the science in the journals that it publishes.

A Counsel Against Despair

One final implication of the convergence principle is that [redacted]. At first, the blur on the screen could represent just about anything. Then, as the slide is focused a bit more, many alternative hypotheses may be ruled out even

though the image cannot be identified unambiguously. Finally, an identification can be made with great confidence. T[redacted].

Thus, [redacted]. Nor is such a situation unique to psychology. It also occurs in more mature sciences. The contradictions may be simply chance occurrences (something we will discuss at length in Chapter 11), or they may be due to subtle methodological differences between experiments. [redacted] (p. 92, Kachka, 2012).

Many other sciences have endured confusing periods of uncertainty before a consensus was achieved (Lewis, 2017; Novella, 2015). Medical science certainly displays this pattern all the time. For example, research into aspirin's role as a cancer preventative has been extremely confusing, uncertain, and nonconverging. Aspirin fights inflammation by inhibiting substances known as cyclooxygenase, or COX, enzymes. Because COX enzymes also are involved in the formation of some cancerous tumors, it was thought that daily aspirin might also inhibit this effect. But actual research on this speculation has produced inconsistent results. Some researchers think that the inconsistency has to do with the fact that the optimal dosage level has not yet been found.

Writer Malcolm Gladwell (2004), in an article titled "The Picture Problem," discusses how people have difficulty understanding why the medical profession still has disagreements about the degree of benefit derived from mammograms (Beck, 2014; Reddy, 2016; University of California, 2016). This is because a mammography picture seems so "concrete" to most people that they think it should be determinative. They fail to understand that human judgment is necessarily involved, and that mammography assessment and disease prediction are inherently probabilistic (Gigerenzer et al., 2007). However, Gladwell, goes on to note that in this area of medicine—just as in psychology—[redacted].

In psychology and many other sciences, [redacted] (Braver et al., 2014; Card, 2011; Schmidt & Oh, [redacted] 2016). In a meta-analysis, the results of several studies that address the same research [redacted]. The effects obtained when one experimental group is compared with another are expressed in a common statistical metric that allows comparison of effects across studies. [redacted]. In some cases, of course, no conclusion can be drawn with confidence, and the result of the meta-analysis is inconclusive.

More and more commentators are calling for a greater emphasis on meta-analysis

as a way of dampening the contentious disputes about conflicting studies in the behavioral sciences. [redacted]. An emphasis on meta-analysis has often revealed that we actually have more stable and useful findings than is apparent from a perusal of the conflicts in our journals.

The National Reading Panel (2000; Ehri et al., 2001) found just this in their meta-analysis of the evidence surrounding several issues in reading education. For example, they concluded that the results of a meta-analysis of the results of 38 different studies indicated "solid support for the conclusion that systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs providing unsystematic or no phonics instruction" (p. 84). In another section of their report, the National Reading Panel reported that a meta-analysis of 52 studies of phonemic awareness training indicated that "teaching children to manipulate the sounds

in language helps them learn to read. Across the various conditions of teaching, testing, and participant characteristics, the effect sizes were all significantly greater than chance and ranged from large to small, with the majority in the moderate range" (p. 5).

[redacted]. It is through meta-analysis that we know that married people are happier than never married people, and that marriage leads to better health outcomes (Myers, 2015, 2017; Robles et al., 2014). It is through meta-analysis that we know that the "brain training" programs advertised on television, radio, and on the web do not work. Although people improve on the specific tasks they train on in these programs, the programs do not improve long-term general cognitive functioning, nor do they have lasting effects on real-world outcomes (Simons et al., 2016). It is through meta-analysis that we know that the personality trait of conscientiousness is related to job performance (Schmidt & Oh, 2016). Many of the studies in the job performance literature on this issue had found nonsignificant results, but when a large number of these studies were amalgamated together, a meta-analysis indicated that there was indeed a modest association. It is through meta-analysis that we know that our ability to predict suicide has not improved in 50 years (Franklin et al., 2017).

[redacted]. In an earlier chapter, we discussed how personal anecdotes had impeded patients and doctors from implementing the recommendation of U.S. Preventive Services Task Force (USPSTF) that the prostate-specific antigen (PSA) test to screen for prostate cancer not be used (Arkes & Gaissmaier, 2012). The USPSTF's review of the scientific evidence indicated that the harms associated with the test (the side effects associated with unnecessary treatment) outweighed the benefits in mortality (which were tiny at best, and perhaps nonexistent). Their recommendation was heavily based on meta-analytic results.

It is likewise in the domain of health psychology. Chida and Hamer (2008) meta-analyzed data from a whopping 281 studies relating the hostility and aggression aspects of the Type A behavior pattern to cardiovascular reactivity (heart rate and blood pressure) in order to establish that there was indeed a relationship. As another example, Currier, Neimeyer, and Berman (2008) meta-analyzed 61 controlled studies of psychotherapeutic interventions for bereaved persons. Their meta-analysis had a disappointing outcome, however. Psychotherapeutic intervention had an immediate effect after bereavement, but had no positive effect at follow-up. The follow-up outcome from this meta-analysis helps to remind us that the outcome of a meta-analysis is not always positive. That is, it does not always tell us that, from a series of widely varying studies, something is there. Just as often it tells us that, when we amalgamate the results from a large number of varying studies—nothing is there!

[redacted]. Also, [redacted]

Summary

In this chapter, we have seen how the breakthrough model of scientific advance is a bad model for psychology and why the gradual-synthesis model provides a better framework for understanding how conclusions are reached in psychology. … Finally, …

Chapter 9

The Misguided Search for the "Magic Bullet": The Issue of Multiple Causation

Learning Objectives

9.1 Explain the concept of interactions between variables in psychological research

9.2 Outline the difficulties in acknowledging multiple causation to a phenomenon

In Chapter 8, we focused on the importance of converging operations and the need to progress to more powerful research methods in order to establish a connection between variables. In this chapter, we go beyond a simple connection between two variables to highlight an important point: …

… ferent variables. To conclude that there is a significant causal connection between variable … For example, researchers have found a negative relationship between amount of television and other media viewing and academic achievement, but they do not claim that the amount of media viewed is the *only* thing that determines academic achievement. That, of course, would be silly, because academic achievement is partially determined by a host of other variables (home environment, quality of schooling, cognitive ability, and the like). In fact, media viewing is only a minor determinant of academic achievement when compared with these other factors. Likewise, Jaffee et al. (2012) examined the literature on

the potential causes of antisocial behavior in youth. The evidence converged on *several* different factors as causal, including: peer deviance, living in a divorced household, parental depression, adolescent motherhood, coercive discipline, and poverty.

[redacted] (Ferguson, 2013; Furuya-Kanamori & Doi, 2016). But [redacted]. [redacted].

Like so many of the other principles discussed in this book, it is important to put the idea of multiple causes in perspective. On the one hand, [redacted]. The world is complicated, and the

determinants of behavior are many and complex. [redacted]. [redacted].

On the other hand, [redacted]. First, [redacted]. Second, [redacted]. Few would argue that a variable that could reduce the number of acts of physical violence by as much as 1 percent annually is not of enormous importance. In short, if the behavior in question is of great importance, then knowing how to control only a small proportion of it can be extremely useful.

There have been medical studies in which a treatment accounted for less than 1 percent of the variability in the outcome, yet the results were considered so startlingly positive that the study was terminated prematurely for ethical considerations—that is, the outcome of the experiment was considered so strong that it was deemed unethical to withhold the treatment from the placebo group (Ferguson, 2009; Rosenthal, 1990).

Likewise, any factor that could cut motor vehicle deaths by just 1 percent would be immensely important—it would save over 250 lives each year. Reducing the homicide rate by just 1 percent would save over 140 lives each year. In short, [redacted]—[redacted].

The Concept of Interaction

[redacted]. This is called [redacted] of interaction: The magnitude of the effect that one variable has may depend on the [redacted]. The concept of interaction really needs to be learned in depth (and with numbers), and statistics texts cover it in detail. We can only mention the concept and give some quick examples here, so this chapter will be a short one.

[redacted] (Evans et al., 2013). [redacted]. [redacted]

[illegible]. For example, researchers might be studying the academic achievement of adolescents to see whether it is a function of life changes such as school transition, pubertal development, residential mobility, and family disruption. It would not be uncommon to find that no single factor had a huge effect, but that when several of these life changes were conjoined together, they resulted in a substantial fall-off in academic achievement.

To understand the logic of what is happening when an interaction such as this occurs, imagine a risk scale where a score of 80–110 represents low risk, 110–125 moderate risk, and 125–150 high risk. Imagine that we had found an average risk score of 82 for children with no stressors, an average risk score of 84 for children with stress factor A, and an average risk score of 86 for children with stress factor B. An interaction effect would be apparent if, when studying children with both risk factor A *and* risk factor B, we found an average risk score of 126. That is, the joint risk when two risk

factors were conjoined was much greater than what would be predicted from studying each risk factor separately.

[illegible]. Grant et al. (2015) found that exposure to a synthetic stress hormone (synthetic glucocorticoids) only had a negative effect on children's cognitive functioning if the children *also* experienced sociodemographic adversity (low maternal education, birth when mother under 18, low income, mother a single parent). The stress hormone did not impair cognitive functioning as long as the child did not experience any sociodemographic adversity.

[illegible]. For example, variations in the so-called *5-HTT* gene have been found to be related to major depression in humans (Hariri & Holmes, 2006). People with one variant (the S allele) are more likely to suffer from major depression than people with the other variant of the gene (the L allele). However, this greater risk for those with the S allele is *only* true for those who have *also* suffered multiple traumatic life events, such as child abuse or neglect, job loss, and/or divorce. [illegible] (Dodge & Rutter, 2011). The relationship between variants of the monoamine oxidase A (*MAOA*) gene and antisocial behavior provides an example. One variant of the gene increases the probability of antisocial behavior, but only if other risk factors are present, such as child abuse, birth complications, or negative home environments (Raine, 2008).

Often it is two psychological characteristics that display an interaction. An example is provided by research on the link between rumination and depression. The tendency to ruminate does predict the duration of depressive symptoms, but it interacts with cognitive styles—rumination predicts lengthened periods of depressive symptoms only when conjoined with negative cognitive styles (Nolen-Hoeksema et al., 2008).

Components of programs may have interactive effects. Developmental psychologist Dan Keating (2007) has reviewed the literature on the consequences of states' Graduated Driver Licensing programs on teen driver safety. These programs work—they lower the rate of teen auto crashes and teen auto fatalities. However, the programs are all different from state to state, each state having somewhat different subsets of several basic components: required driver education, passenger restrictions, night driving restrictions, extended legal age, minimum practice driving requirements, and extended learner's permit time. Thus, the question becomes whether each of these components is causally effective and whether they have any interactive effects. Research indicates that no *one* of the components lowers teen crash or fatality rates. However, in *combination* they can lower the number of teen fatalities by over 20 percent.

Thus, [illegible]. First, [illegible]. Second, [illegible]

[redacted].

Clinical psychologist Scott Lilienfeld (2006) discusses the continuum of causal influence for variables—from strong to weak. [redacted]. Finally, [redacted].

The Temptation of the Single-Cause Explanation

It seems that the basic idea that complex events in the world are multiply determined should be an easy one to grasp. In fact, the concept *is* easy to grasp and to apply when the issues are not controversial. However, when our old nemesis, preexisting bias (see Chapter 3), rears its head people have a tendency to ignore the principle of multiple causation. How many times do we hear people arguing about such emotionally charged issues as the causes of crime, the distribution of wealth, the causes of poverty, changes in marriage rates, and the effect of capital punishment in a way that implies that these issues are simple and unidimensional and that outcomes in these areas have a single cause? These examples make it clear that people will sometimes acknowledge the existence of multiple causes if asked *directly* about multiple causes; but seldom will they *spontaneously* offer many different causes as an explanation for something they care about. Most often, people adopt a "zero sum" attitude toward potential causes—that all causes compete with one another and that emphasizing one necessarily reduces the emphasis on another.

Under emotional influence, we tend to forget the principle of multiple causation. For example, consider discussions of crime by people on opposite ends of the political spectrum. Liberals may argue that people of low-socioeconomic status who commit crimes may themselves be victims of their circumstances (e.g., joblessness, poor housing, poor education, and lack of hope about the future). Conservatives may reply that a lot of poor people do not commit crimes; therefore, economic conditions are not the cause. Instead, the conservative may argue, it is personal values and personal character that determine criminal behavior. Neither side in the debate ever seems to acknowledge that *both* individual factors and environmental factors contribute to criminal behavior. [redacted].

Consider also discussions of the causes of complex economic outcomes. These outcomes are hard to predict precisely because they are multiply determined. For example, economic debate has focused on a problem of the last several decades with important social implications: [redacted] (Brooks, 2008; Caldwell, 2016; Conard, 2016; Fairless, 2017; Lemann, 2012; Murray, 2012). [redacted]

[illegible] What have economic studies found with respect to these many alternative causes? You guessed it. [illegible]. It also appears that [illegible]. For example, [illegible].

It can be particularly difficult to acknowledge multiple causation when the different potential causes are attached to different ideological positions. Then, it becomes very tempting for each side to promote its own cause (or causes) and

dismiss the causes of their ideological opponents. The causes of poverty provide a prime example of this tendency—with liberals and conservatives having different favorite causal models and tending to denigrate the causal models of the other side. The answer to this impasse is for each side to acknowledge multiple causation at the outset and to concede that some of the multiple causes probably include those of their ideological opponents.

Psychologist Jonathan Haidt of New York University participated in an endeavor that attempted to break this impasse by truly acknowledging multiple causation (AEI/Brookings Working Group on Poverty, 2015). A group of experts from across the ideological spectrum agreed to come up with a set of solutions for poverty that were consensus positions—endorsed by both sides of the ideological divide. Striving for consensus necessarily meant that members of the panel would have to endorse solutions to poverty that came from their ideological opponents. The group was largely successful in producing a report that truly acknowledged multiple causation when it came to their policy recommendations. For example, the liberals on the panel agreed that the data supported the conservative policy recommendation of promoting a new cultural norm surrounding parenthood and marriage, as well as the policy recommendation of promoting delayed, responsible childbearing. Likewise, the conservatives on the panel agreed that the data supported the liberal policy recommendation of making work pay better for the less educated and for ensuring that jobs were available. The consensus recommendations thus embodied the principle of multiple causation.

Like economic problems, [illegible]. Take the problem of learning disabilities, for example, which educational psychologists, cognitive psychologists, and developmental psychologists have investigated extensively. [illegible] (Peterson & Pennington, 2012; Seidenberg, 2017; Tanaka et al., 2011). [illegible] (Peterson & Pennington, 2012). [illegible]. [illegible]. The reason it would be wrong is that [illegible] ronments (Cunningham & Zibulsky, 2014; Seidenberg, 2017). [illegible].

A similar situation characterizes the causes and treatment of depression. [illegible]. Likewise, a multiplicity of treatments combined—medication plus psychotherapy—seems to result in the best therapeutic outcome (Engel, 2008).

Once the multiple causes of a complex phenomenon are found, if the phenom-

enon is a problem, this necessarily means that the solution to the problem will require multiple interventions. Decades ago we had a major health problem—an epidemic of smoking, a habit linked to many diseases. In recent decades, various interventions have reduced the level of smoking in our society: Tobacco advertising was banned, tobacco taxes were raised, the nicotine patch became available, smoking was banned in public places, and many more interventions were instituted (Brody, 2011). Slowly, over decades, the rate of smoking went down because of these multiple interventions targeted at its many causes.

Just as it took many different interventions to reduce smoking years ago, it will take multiple societal interventions to halt and reverse our current national epidemic of obesity (King, 2013; Taubes, 2017). The reason is that our current obesity epidemic started a couple of decades ago because of many different trends coinciding: Decreased walking, fewer meals were prepared at home when more women entered

the workforce, the fast-food industry exploded in size, food advertising became ubiquitous, electronic entertainment made children sedentary, portions increased in size, along with many other factors (Hewer, 2014; University of California, 2015b). The solution to this national problem will have to be correspondingly multifaceted. The Wellness Newsletter of the University of California warns that "it's overly simplistic to blame the obesity epidemic solely on people eating too much because of lack of willpower and sedentary lifestyles. If there ever was a multifactorial condition, obesity is it—a complex of interacting genetic, metabolic, behavioral, hormonal, psychological, cultural, environmental, and socioeconomic factors" (p. 1, University of California, 2015b). Science writer Gina Kolata (2016b) puts it even simpler by titling her article on obesity: No Single Answer.

Summary

The single lesson of this chapter is an easy but important one. [redacted]

Chapter 10

The Achilles' Heel of Human Cognition: Probabilistic Reasoning

Learning Objectives

10.1 Describe how people use “person-who” arguments to refute psychological findings

10.2 Explain how probabilistic prediction means learning to live with some uncertainty

10.3 Outline some pitfalls in dealing with probability, including the gambler’s fallacy and ignoring sample size

Question:
Men are taller than women, right?

Answer:
“Right.”

Question:
All men are taller than all women, right?

Answer:
“Wrong.”

Correct. Believe it or not, we are going to devote part of this chapter to something that you just demonstrated you knew by answering the previous two questions. But don’t skip the chapter just yet, because there are some surprises waiting in the explanation

of what seems like a very simple principle.

[redacted]. You correctly interpreted the statement as reflecting a probabilistic trend rather than a fact that holds in every single instance. [redacted]. That is, [redacted]. [redacted]: It tends to be warmer near the equator. Families tend to have fewer than eight children. Most parts of the earth tend to have more insects than humans. These are all statistically demonstrable trends, yet there are exceptions to every one of them. [redacted].

Before his death of lung cancer, neurosurgeon Paul Kalanithi (2016) wrote a moving book about dealing with the disease in the latter days of his life. In his book, he discussed how doctors present prognoses to patients, and was particularly harsh on physicians who did not emphasize to patients that prognoses were probabilistic. He recommended presenting patients with intervals ("Most patients live many months to a couple of years") rather than specific best guesses ("Median survival is eleven months"). He felt that phrases like "most patients live many months to a couple of years" more fundamentally portrayed the probabilistic nature of prediction.

Americans received a sad lesson in the probabilistic nature of medical knowledge in the summer of 2008 when much-loved political broadcaster Tim Russert died of a heart attack at age 58. Russert took cholesterol pills and low-dose aspirin, rode an exercise bike, and had yearly stress tests, yet he still died early of a heart attack. The fact that he had been fairly vigilant toward his health led many readers of the *New York Times* to write in saying that the doctors must have missed something. These readers did not understand that medical knowledge is probabilistic. Every failure to predict is not a mistake. In fact, his doctors missed nothing. They applied their probabilistic knowledge as best they could—but this does not mean that they could predict individual cases of heart attack. Science writer Denise Grady (2008) tells us that, based on his stress test and many other state-of-the-art diagnostics that Mr. Russert was given in his last exam, the doctors estimated—from a widely used formula—that Mr. Russert's probability of a heart attack in the next ten years was 5 percent. This means that 95 out of 100 people with Mr. Russert's medical profile should not have a heart attack in the next ten years. Mr. Russert was just one of the unlucky five—and medical science, being probabilistic, cannot tell us in advance who those unlucky five will be.

The Tim Russert example provides an opportunity to emphasize that [redacted]. Here is what we mean by this: [redacted]. [redacted]. But after they are dead, those five people most definitely *do* have names. For example, Tim Russert turned out to be one of the five. He is no less dead than he would be if we could have named him in advance. We must get over this feeling that, because of its numerical abstraction, probabilistic prediction is not real.

[redacted]. Recall from Chapter 8 the point that, because of cell phone talking and texting in cars, hundreds of Americans will die unnecessarily in crashes in the upcoming year. Because this is a probabilistic prediction, I cannot tell you who these Americans will be. However, the prediction is no less real just because it is probabilistic.

[redacted]. Science writer Natalie Angier (2007) discusses

how some people think that seismologists really *can* predict individual earthquakes, but that they do not make these predictions public so as "not to create a panic." One seismologist received a letter from a woman asking him to tell her if he ever sent his children to see out-of-town relatives. From this example, Angier notes that people seem to prefer to believe that authorities are engaged in monstrous lying than to simply admit that there is uncertainty in science.

Political pollsters, for example, learn to live with this uncertainty, even though the public that they serve is not comfortable with it. After the 2016 presidential election in the United States, pollsters took a lot of flack for their faulty predictions. Actually though, the pollsters were pretty close in calling the popular vote. What they failed to predict correctly was the outcome in the electoral college. Pollster and statistician Nate Silver was particularly victimized by public misunderstanding of probabilistic prediction. Close to the election, his prediction was a 71 percent probability that Hillary Clinton would win the electoral college. Democrats were furious with him because

most other pollsters were setting Clinton's chance of winning the electoral college at over 90 percent (Flint & Albert, 2016; Hemingway, 2016; Lohr & Singer, 2016). One Princeton poll had the probability of a Clinton electoral college win at 99 percent! Democratic websites accused Silver of skewing his analysis in Donald Trump's favor. Of course, after the election, Silver got little credit, because he had still predicted the wrong winner. He received little praise from the public for providing a probabilistic estimate that more accurately reflected the uncertainty in the race.

Virtually all the facts and relationships that have been uncovered by the science of psychology are stated in terms of probabilities. There is nothing unique about this. Many of the laws and relationships in other sciences are stated in probabilities rather than certainties. The entire subdiscipline of population genetics, for example, is based on probabilistic relationships. Physicists tell us that the distribution of the electron's charge in an atom is described by a probabilistic function. Thus, the fact that behavioral relationships are stated in probabilistic form does not distinguish them from those in other sciences.

"Person-Who" Statistics

[illegible]. Smoking causes lung cancer and a host of other health problems. Voluminous medical evidence documents this fact (Gigerenzer et al., 2007). Yet will everyone who smokes get lung cancer, and will everyone who refrains from smoking be free of lung cancer? Most people know that these implications do not follow. The relationship is probabilistic. [illegible]. [illegible]. It cannot tell us which ones will die, though. [illegible]. We are all aware of this—or are we? How often have we seen a nonsmoker trying to convince a smoker to stop by citing the smoking—lung-cancer statistics, only to have the smoker come back with "Oh, get outta here! Look at old Joe Ferguson down at the store. Three packs of Camels a day since he was sixteen! Ninety-one years old and he looks great!" The obvious inference that one is supposed to draw is that this single case somehow invalidates the relationship.

It is surprising and distressing how often this ploy works. [illegible]. [illegible]. If people think a single example can invalidate a law, they must feel the law should hold in every case. In short, they have failed to understand the law's probabilistic nature. There will always be a "person who" goes against even the strongest of trends.

Psychologists call instances like the "old Joe Ferguson" story examples of the use of

. The ubiquitous "person who" is usually trotted out when we are confronted with hard statistical evidence that contradicts a previously held belief. Thus, it could be argued that

ly use the "person who" as a technique to invalidate facts

. However, the work of psychologists who have studied human decision making and reasoning suggests that the tendency to use the "person who" comes not simply from its usefulness as a debating strategy. Instead,

. It is so much of an Achilles' heel that probabilistic reasoning is at the heart of the operational definition of human rationality (Stanovich, West, & Toplak, 2016).

Probabilistic Reasoning and the Misunderstanding of Psychology

The findings of psychology are often misunderstood because of the problems people have in dealing with probabilistic information.

. Most people understand the statement "smoking causes lung cancer" in the same way (although old "Joe Ferguson" can be convincing to some smokers who do not want to believe that their habit may be killing them!). However,

Most psychology instructors have witnessed a very common reaction when they discuss the evidence on certain behavioral relationships. For example, the instructor may present the fact that children's scholastic achievement is related to the socioeconomic status of their households and to the educational level of their parents. This statement often prompts at least one student to object that he has a friend who is a National Merit Scholar and whose father finished only eighth grade. Even those who understood the smoking—lung-cancer example tend to waver at this point.

Most people understand that many treatments, theories, and facts developed by medical science are probabilistic. They understand that, for example, a majority of patients, but not all of them, will respond to a certain drug. Medical science, however, often cannot tell in advance which patients will respond. Often all that can be said is that if 100 patients take treatment A and 100 patients do not, after a certain period the 100 patients who took treatment A will *collectively* be better off. I mentioned in an earlier chapter that I take a medication called Imitrex (sumatriptan succinate) for relief from migraine headaches. The information sheet accompanying this drug tells me that controlled studies have demonstrated that, at a particular dosage level, 57 percent of patients taking this medication receive relief in two hours. I am one of the lucky 57 percent—but neither the drug company nor my physician could give me a guarantee that I would not be one of the unlucky 43 percent. The drug does not work in every case.

s.

—

[illegible]. [illegible] (we will discuss actuarial prediction in more detail in the next chapter).

Consider an unhealthy person going to a physician. The person is told that unless he or she exercises and changes diet, he or she has a high risk of heart attack. We are not tempted to say that the doctor has no useful knowledge because he or she cannot tell the person that without a change of diet he or she will have a heart attack on September 18, 2024. We tend to understand that the physician's predictions are probabilistic and cannot be given with that level of precision. It is likewise when geologists tell us that there is a 60 percent probability of a magnitude 7.0 or greater earthquake in a certain area in the next 30 years (Silver, 2012). We do not denigrate their knowledge because they cannot say that there will be an earthquake exactly *here* on July 5, 2023.

Yet not everyone fully understands this. In April of 2009, an earthquake occurred in L'Aquila, Italy, and killed 309 people (Diacu, 2012; Radford, 2016b). It injured over 1,500 people. Incredibly, in 2012, an Italian court levied a criminal conviction against six of the country's seismologists for not accurately predicting the earthquake! The conviction was overturned in 2016, but it demonstrates how difficult it is for the public (and sometimes the courts) to understand the fundamental idea that probabilistic prediction does not allow perfect prediction in individual cases (Silver, 2012).

[illegible]—[illegible]. It is likewise when a clinical psychologist recommends a program for a child with self-injurious behavior. The psychologist judges that there is a higher probability of a good outcome if a certain approach is followed. But unlike the heart attack and earthquake examples, the psychologist is often confronted with questions like "but *when* will my child be reading at grade level?" or "exactly how long will he have to be in this program?" These are unanswerable questions—in the same way that the questions about exactly when the earthquake or the heart attack will occur are also unanswerable questions. They are unanswerable because in all these cases—the heart attack, the learning disabled child, the earthquake, the child with self-injurious behavior—the prediction being made is probabilistic.

For these reasons, [illegible]. There is a profound irony here. [illegible] Yet, [illegible].

Psychological Research on Probabilistic Reasoning

In the past three decades, the research of psychologists such as Daniel Kahneman of Princeton University (winner of the Nobel Prize in 2002) and the late Amos Tversky has revolutionized the way we think about people's reasoning abilities (Lewis, 2017). In the course of their studies, these investigators have uncovered some fundamental principles of probabilistic reasoning that are absent or, more commonly, insufficiently developed in many people. As has often been pointed out, it should not be surprising that they are insufficiently developed. As a branch of mathematics, probability theory is a very recent development. The key initial developments did not occur until the sixteenth and seventeenth centuries (Hand, 2014; Mazur, 2016), and many essential

developments date not much past the twentieth century.

The dates of the initial developments in probability theory highlight a significant fact: Games of chance existed centuries before the fundamental laws of probability were discovered. Here is another example of how personal experience does not seem to be sufficient to lead to a fundamental understanding of the world (see Chapter 7). It took formal study of the laws of probability to reveal how games of chance work. Thousands of gamblers and their "personal experiences" throughout history were insufficient to uncover the underlying nature of games of chance.

The problem is that [redacted]. [redacted].

"Why did they raise my insurance rate," you might wonder, "and why is John's rate higher than Bill's? Is Social Security going broke? Is our state lottery crooked? Is

crime increasing or decreasing? Why do doctors order all those tests? Why can people be treated with certain rare drugs in Europe and not in the United States? Do women really make less than men in comparable jobs? Do international trade deals cost Americans jobs and drive down wages? Is educational achievement in Japan really higher than here? These are all good questions—concrete, practical questions about our society and how it works. To understand the answers to each of them, however, one must think statistically.

Clearly, a complete discussion of statistical and probabilistic thinking is beyond the scope of this book. We will, however, briefly discuss some of the more common pitfalls of probabilistic reasoning. A good way to start developing the skill of probabilistic thinking is to become aware of the most common fallacies that arise when people reason statistically.

Insufficient Use of Probabilistic Information

One finding that has been much replicated is that [redacted] (the vividness problem discussed in Chapter 4). [redacted]. Here is a problem (see Stanovich, 2010) that even experienced decision makers such as physicians find difficult: Imagine that the virus that causes AIDS (HIV) occurs in 1 in every 1,000 people. Imagine also that there is a test to diagnose the disease that always indicates correctly that a person who has HIV actually has it. Finally, imagine that the test has a false-positive rate of 5 percent. This means that the test wrongly indicates that HIV is present in 5 percent of the cases in which the person does not have the virus. Imagine that we choose a person randomly and administer the test and that it yields a positive result (indicates that the person is HIV-positive). What is the probability that the individual actually has the HIV virus, assuming that we know nothing else about the individual's personal or medical history?

The most common answer to this problem (even among experienced physicians) is 95 percent. The correct answer is approximately 2 percent. People vastly overestimated the probability that a positive result truly indicated the disease because of the tendency to overweight the case information and underweight the base rate information (that only 1 in 1,000 people are HIV-positive). A little logical reasoning can help to illustrate the profound effect that base rates have on probabilities. Of 1,000 people, only 1 will actually be HIV-positive. If the other 999 (who do not have the disease) are tested, the test will indicate incorrectly that approximately 50 of them have the virus (0.05 multiplied by 999) because of the 5 percent false-positive rate. Thus, of the 51 patients testing positive, only 1 (approximately 2 percent) will actually be HIV-

positive. In short, the base rate is such that the vast majority of people do not have the virus (only 1 in 1,000). This fact, combined with a substantial false-positive rate, ensures that, in absolute numbers, the vast majority of positive tests will be of people who do not have the virus.

Although most people recognize the correctness of this logic, their initial tendency is to discount the base rates and overweight the clinical evidence. In short, [illegible] actually know better but are initially drawn to an incorrect conclusion. Psychologists [illegible] (Pohl, 2017). [illegible].

In this problem, the case evidence (the laboratory test result) seems tangible and concrete to most people, whereas the probabilistic evidence seems, well—probabilistic. This reasoning, of course, is fallacious because [illegible]. A clinical test misidentifies the presence of a disease with a certain *probability.* The situation is one in which two probabilities—the probable diagnosticity of the case

evidence and the prior probability (base rate)—must be combined if one is to arrive at a correct decision. There are right and wrong ways of combining these probabilities, and more often than not—particularly when the case evidence gives the illusion of concreteness (recall our discussion of the vividness problem in Chapter 4)—people combine the information in the wrong way.

The HIV example above also illustrates the importance of paying attention to the false-positive rate when interpreting test results. In that example, a substantial false-positive rate (5 percent) combined with a low base rate for the disease (only 1 in 1,000) resulted in the following consequence: More people with a *positive* test result did *not* have the disease than did have it. Attention to false-positives is a critical concern in all diagnostic testing, including in medicine where, despite great advances in treatment and diagnosis, most clinical tests still have substantial false-positive rates. In one study of 30,000 older men, it was found that after taking four screening tests for prostate, lung, and colorectal cancer, more than *one-third* of the men received a false-positive result—the test indicated that they had cancer when in fact they were cancer free (Croswell et al., 2009).

Failure to Use Sample-Size Information

Consider these two problems (see Kahneman, 2011):

1. A certain town is served by two hospitals. In the larger hospital, about 45 babies are born each day, and in the smaller hospital, about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it is higher than 50 percent, sometimes lower. For a period of one year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?
   a. The larger hospital
   b. The smaller hospital
   c. About the same
2. Imagine an urn filled with balls, two-thirds of which are of one color and one-third of which are of another. One individual has drawn 5 balls from the urn and found that 4 are red and 1 is white. Another individual has drawn 20 balls and found that 12 are red and 8 are white. Which of the two individuals should feel more confident that the urn contains two-thirds red balls and one-third white balls, rather than vice versa? What odds should each individual give?

In problem 1, the majority of people answer "about the same." People not choosing this alternative pick the larger and the smaller hospital with about equal frequency.

Because the correct answer is the smaller hospital, approximately 75 percent of subjects given this problem answer incorrectly. These incorrect answers result from an inability to recognize the importance of sample size in the problem. Other things being equal, a larger sample size always more accurately estimates a population value. Thus, on any given day, the larger hospital, with its larger sample size, will tend to have a proportion of births closer to 50 percent. Conversely, a small sample size is always more likely to deviate from the population value. Thus, the smaller hospital will have more days on which the proportion of births displays a large discrepancy from the population value (60 percent boys, 40 percent boys, 80 percent boys, etc.).

In problem 2, most people feel that the sample of 5 balls provides more convincing evidence that the urn is predominantly red. Actually, the probabilities are in the opposite direction. The odds are 8 to 1 that the urn is predominantly red for the 5-ball sample, but they are 16 to 1 that the urn is predominantly red for the 20-ball sample. Even though the proportion of red balls is higher in the 5-ball sample (80 percent versus 60 percent), this is more than compensated for by the fact that the other sample is four

times as large and, thus, is more likely to be an accurate estimate of the proportions in the urn. The judgment of most subjects, however, is dominated by the higher proportion of red in the 5-ball sample and does not take adequate account of the greater reliability of the 20-ball sample.

[illegible] Smaller samples will always generate more extreme values. Psychologist Daniel Kahneman (2011) shows an example of how the failure to apply this principle can send us on a wild goose chase in search of causal theories when none are needed. He pointed out that a study of 3,141 counties in the United States found that the counties in which the incidence of kidney cancer was lowest tended to be rural counties that were sparsely populated. Kahneman (2011) pointed out how easy it would be to come up with a causal theory about why this was the case: "the clean living of the rural lifestyle—no air pollution, no water pollution, access to fresh food without additives" (p. 109). The only problem with this causal theory is that it does not account for another finding from the *same* study: The counties in which the incidence of kidney cancer was *highest* tended to be rural counties that were sparsely populated! Had we been told this last fact first, we might have started to posit explanations of rural counties having more smoking, drinking, and high-fat diets. But this, and the earlier explanation for the low-incidence counties, would both have been off the mark. What we have here is the hospital problem discussed previously playing out in real life. Rural counties with sparse populations are small samples, and they are bound to produce more extreme values of *all* types—extremely high values and extremely low values.

Many people have problems recognizing that they are in situations involving sampling. That is, they have difficulty realizing that they are dealing with a sample rather than the entire entity. Failure to realize this leads them to miss the fact that [illegible]. For example, when a blood test is ordered by your physician, what is taken from you will be a *sample* and *it* will be assessed, not the state of your entire blood system. [illegible]. [illegible], because the cells in the sample and their composition and properties will necessarily deviate a little bit from absolute truth because the test cannot measure your entire blood system. In short, your physician is making assumptions about your entire composition from a tiny sample.

It is likewise when a tumor is biopsied. There is some error involved, because the biopsy yields only a small sample from a larger tumor. Medical writer Tara Parker-Pope (2011), in discussing the biopsy done for suspected prostate cancer, informs us that a very common type of biopsy samples only about one three-thousandth of the

prostate. She cites evidence that staging and grading mistakes occur in about 20 percent of specimens. The point to realize is that it is the same when we are measuring behavior. We often are taking a small sample to represent a much larger population of behavior.

The Gambler's Fallacy

Please answer the following two problems:

Problem A: Imagine that we are tossing a fair coin (a coin that has a 50/50 chance of coming up heads or tails) and it has just come up heads five times in a row. For the sixth toss, do you think that

________ It is more likely that tails will come up than heads?

________ It is more likely that heads will come up than tails?

________ [illegible]?

Problem B: When playing slot machines, people win something one out of every 10 times. Julie, however, has just won on her first four plays. What are her chances of winning the next time she plays? ________ out of ________

These two problems probe whether a person is prone to the so-called [illegible]. [illegible]. Most games of chance that use proper equipment have this property. For example, [illegible]. Half the numbers on a roulette wheel are red, and half are black (for purposes of simplification, we will ignore the green zero and double zero), so the odds are even (0.50) that any given spin will come up red. Yet after five or six consecutive reds, many bettors switch to black, thinking that it is now more likely to come up. This is [illegible]. In this case, the bettors are wrong in their belief. The roulette wheel has no memory of what has happened previously. Even if 15 reds in a row come up, the probability of red coming up on the next spin is still 0.50.

In problem A, some people think that it is more likely that either heads or tails will come up after five heads, and they are displaying the gambler's fallacy by thinking so. [illegible].

The gambler's fallacy is not restricted to the inexperienced. Research has shown that even habitual gamblers, who play games of chance over 20 hours a week, still display belief in the gambler's fallacy (Petry, 2005). In fact, [illegible] (Toplak et al., 2007).

It is important to realize that [illegible] is not restricted to games of chance. It [illegible]. The genetic makeup of babies is an example. Psychologists, physicians, and marriage counselors often see couples who, after having two female children, are planning a third child because "We want a boy, and it's *bound* to be a boy this time." This, of course, is the gambler's fallacy. The probability of having a boy (approximately 50 percent) is exactly the same after having two girls as it was in the beginning. The two previous girls make it *no more likely* that the third baby will be a boy.

The gambler's fallacy stems from many mistaken beliefs about probability. One

is the belief that if a process is truly random, no sequence—not even a small one (six coin flips, for instance)—should display runs or patterns. People routinely underestimate the likelihood of runs (HHHH) and patterns (HHTTHHTTHHTT) in a random sequence. For this reason, people cannot generate truly random sequences when they try to do so. The sequences that they generate tend to have too few runs and patterns. When generating such sequences, people alternate their choices too much in a mistaken effort to destroy any structure that might appear (Fischer & Savranevski, 2015; Scholl & Greifeneder, 2011).

Those who claim to have psychic powers can easily exploit this tendency. Consider a demonstration sometimes conducted in college psychology classes. A student is told to prepare a list of 200 numbers by randomly choosing from the numbers 1, 2, and 3 over and over again. After it is completed, the list of numbers is kept out of view of the instructor. The student is now told to concentrate on the first number on the list, and the instructor tries to guess what the number is. After the instructor guesses, the student tells the class and the instructor the correct choice. A record is kept of whether

the instructor's guess matched, and the process continues until the complete record of 200 matches and nonmatches is recorded. Before the procedure begins, the instructor announces that she or he will demonstrate "psychic powers" by reading the subject's mind during the experiment. The class is asked what level of performance—that is, percentage of "hits"—would constitute empirically solid evidence of psychic powers. Usually a student who has taken a statistics course volunteers that, because a result of 33 percent hits could be expected purely on the basis of chance, the instructor would have to achieve a larger proportion than this, probably at least 40 percent, before one should believe that she or he has psychic powers. The class usually understands and agrees with this argument. The demonstration is then conducted, and a result of more than 40 percent hits is obtained, to the surprise of many.

The students then learn some lessons about randomness and about how easy it is to fake psychic powers. The instructor in this example merely takes advantage of the fact that people do not generate enough runs: They alternate too much when producing "random" numbers. In a truly random sequence of numbers, what should the probability of a 2 be after three consecutive 2s? One-third, the same as the probability of a 1 or a 3. But this is not how most people generate such numbers. After even a small run, they tend to alternate numbers in order to produce a representative sequence. Thus, on each trial in our example, the instructor merely picks one of the two numbers that the student did not pick on the previous trial. Thus, if on the previous trial the student generated a 2, the instructor picks a 1 or a 3 for the next trial. If on the previous trial the subject generated a 3, the instructor picks a 1 or a 2 on the next trial. This simple procedure usually ensures a percentage of hits greater than 33 percent—greater than chance accuracy without a hint of psychic power.

[illegible] was illustrated humorously in the controversy over the iPod's "shuffle" feature that broke out in 2005 (Levy, 2005; Froelich et al., 2009). This feature plays the songs loaded into the iPod in a random sequence. Of course, knowing the research I have just discussed, many psychologists and statisticians chuckled to themselves when the inevitable happened—users complained that the shuffle feature could not be random because they often experienced sequences of songs from the same album or genre (Ziegler & Garfield, 2012). Technical writer Steven Levy (2005) described how he had experienced the same thing. His iPod seemed always to have a fondness for Steely Dan in the first hour of play! But Levy was smart enough to accept what the experts told him: [illegible].

A Further Word About Statistics and Probability

A Further Word About Statistics and Probability

These, then, are just a few of the shortcomings in statistical reasoning that obscure an understanding of psychology. More complete and detailed coverage is provided in Kahneman's *Thinking, Fast and Slow* (2011). Introductions to many of these ideas (and good places to start for those who lack extensive statistical training) are contained in Hastie and Dawes's *Rational Choice in an Uncertain World* (2010), Baron's *Thinking and Deciding* (2008), my own *Decision Making and Rationality in the Modern World* (2010), Charles Wheelan's *Naked Statistics: Stripping the Dread from Data* (2013), and Jordan Ellenberg's *How Not to Be Wrong: The Power of Mathematical Thinking* (2014).

The probabilistic thinking skills discussed in this chapter are of tremendous practical significance. Because of inadequately developed probabilistic thinking abilities, physicians choose less effective medical treatments (Croskerry, 2013); people fail to assess accurately the risks in their environment (Fischhoff & Kadvany, 2011); information is misused in legal proceedings (Gigerenzer et al., 2007); unnecessary

surgery is performed (Gigerenzer et al., 2007; Groopman, 2007); and costly financial misjudgments are made (Lewis, 2017; Thaler, 2015; Zwieg, 2008).

Of course, a comprehensive discussion of statistical reasoning cannot be carried out in a single chapter. Our goal was much more modest: [illegible]. Unfortunately, there is no simple rule to follow when confronted with statistical information. Unlike some of the other components of scientific thinking that are more easily acquired, [illegible].

[illegible]: Such is the case with statistics and psychology. [illegible] (Evans, 2015). A past president of the Association for Psychological Science, Morton Ann Gernsbacher (2007), derived a list of 10 things of intellectual value that she thinks psychological training specifically instills, and 4 of her 10 were in the domains of statistics and methodology. [illegible] (Parry, 2012).

Ludy Benjamin, winner of a prestigious APA teaching award, discussed the most important features that he says should be in an introductory psychology class. While acknowledging that of course such a class must present the most important findings in the discipline, Benjamin went on to say that he thought that "in the long run, [illegible]. [illegible]" (Dingfelder, 2007, p. 26).

This legacy is much valued in the real world outside of the academic psychology department. *Money Magazine* listed the 21 most valuable career skills in their survey of business and industry (Weisser et al., 2016) and the list was chock-full of statistics and data-analysis skills (data mining, forecasting, facility in statistical software, data modeling, etc.). And if you are a psychology major reading this, you should note that there are more and more CEOs out there like Boaz Salik (2016). In interviews, he asks job applicants to his consulting firm how to calculate the joint probability of two events! The question should be a piece of cake for any psychology major, but might

events. The question should be a piece of cake for any psychology major, but might trip up anyone not comfortable with statistics and probability.

Our current world is awash in statistics and graphic displays of numbers. In medicine, finance, advertisements, and on the news, we are presented with claims based on statistics (Silver, 2012). We need to learn to evaluate them, and fortunately [redacted] Clearly, one of the goals of this book is to make research in the discipline of psychology more accessible to the general reader. However, [redacted] (as is the case in many other fields, such as economics, sociology, and genetics) [redacted]. Thus, although this chapter has served as an extremely brief lesson in statistical thinking, its main purpose has been to highlight the existence of an area of expertise that is critical to a full understanding of psychology.

Summary

As in most sciences, [redacted]— [redacted]. [redacted] (just as is the case in other sciences). One thing that prevents the understanding of much psychological research is that many people have difficulty thinking in probabilistic terms. In this chapter, we discussed several well-researched examples of how probabilistic reasoning goes astray for many people: They make insufficient use of probabilistic information when they also have vivid testimonial evidence available; they fail to take into account the fact that larger samples give more accurate estimates of population values; and, finally, they display the gambler's fallacy (the tendency to see links among events that are really independent). The gambler's fallacy derives from a more general tendency that we will discuss in the next chapter: the tendency to fail to recognize the role of chance in determining outcomes.

Chapter 11

The Role of Chance in Psychology

Learning Objectives

11.1 Explain why chance impedes interpretation of scientific evidence

11.2 Explain how people misunderstand the meaning of coincidence

11.3 Differentiate between actuarial and clinical predictions

In the last chapter, we discussed the importance of probabilistic trends, probabilistic thinking, and statistical reasoning. In this chapter, we will continue that discussion with an emphasis on the difficulties of understanding the concepts of randomness and chance. We will emphasize how people often misunderstand the contribution of research to clinical practice because of a failure to appreciate how thoroughly the concept of chance is integrated within psychological theory.

The Tendency to Try to Explain Chance Events

[illegible]. [illegible]. This strong tendency to search for structure has been studied by psychologists. [illegible].

[illegible]

Nevertheless, [illegible]. The quest for conceptual understanding is maladaptive when it takes place in an environment in which there is nothing to conceptualize. What plays havoc with one of the most distinguishing features of human cognition? What confounds our quest for structure and obscures understanding? You guessed it: probability. Or, more specifically, chance and randomness.

[illegible] Again, recall a previous example: Smoking causes lung cancer. A systematic, explainable aspect of biology links smoking to this particular disease. But not all smokers contract lung cancer. The trend is probabilistic. Perhaps we will eventually be able to explain why some smokers do not contract cancer. However, for the time being, this variability must be ascribed to the multitude of chance factors that determine whether a person will contract a particular disease.

As this example illustrates, [illegible]. A coin toss is a chance event, but not because it is in principle impossible to determine the outcome by measuring the angle of the toss, the precise composition of the coin, and many other variables. In fact, the outcome of a toss is determined by all these variables. But [illegible]. [illegible].

[illegible]. Often, however, [illegible]. Psychologists have conducted experiments on this phenomenon. In one experimental situation, subjects view a series of stimuli that vary in many different dimensions. The subjects are told that some stimuli belong to one class and other stimuli belong to another. Their task is to guess which class each of a succession of stimuli belongs to. However, the researcher actually assigns the stimuli to classes randomly. Thus, there is no rule except randomness. The subjects, however, rarely venture randomness as a guess. Instead, they often concoct extremely elaborate and complicated theories to explain how the stimuli are being assigned.

The thinking of many financial analysts illustrates how difficult it is to acknowledge the large effect of randomness in certain domains. It is common for financial analysts to concoct elaborate explanations for every little fluctuation in stock market prices. In fact, much of this variability is simply random fluctuation (Ellis, 2016; Kahneman, 2011). What we should be hearing many nights on television is something like "The Dow Jones average gained 27 points today because of random fluctuation in a complex interacting system." You will never hear this headline, because financial analysts want to imply that they can explain *everything*—every little burp in market behavior. They continue to imply to their customers (and perhaps themselves believe) that they can "beat the market" when there is voluminous evidence that the vast majority of them can do no such thing. Throughout most of the last several decades, if you had bought all of the 500 stocks in the Standard and Poor's Index and simply held them (what we might call a no-brain strategy—a strategy you could actually carry out by buying a mutual fund that tracks that index), then you would have had higher returns than over three quarters of the money managers on Wall Street (Bogle, 2015; Ellis, 2016; Investor's Guide, 2017; Malkiel, 2016). You would also have beaten 80 percent of the financial newsletters that subscribers buy at rates of up to $1,000 per year.

But what about the managers who *do* beat the no-brain strategy? You might be

wondering whether this means that they have some special skill. We can answer that question by considering the following thought experiment. One hundred monkeys have each been given ten darts, and they are each going to throw them at a wall containing the names of each of the Standard and Poor's 500 stocks. Where the darts land will define that monkey's stock picks for the year. How will they do a year later? How many will beat the Standard and Poor's 500 Index? You guessed it. Roughly half of the monkeys. Would you be interested in paying the 50 percent of the monkeys who beat the index a commission to make your picks for you next year?

The logic by which purely random sequences seem to be the result of predictable factors is illustrated by a continuation of this example of financial predictions. Imagine that a letter comes in the mail informing you of the existence of a stock-market-prediction newsletter. The newsletter does not ask for money but simply tells you to test it out. It tells you that IBM stock is going to go up during the next month. You put the letter away, but you do notice IBM stock does go up the next month. Having read a book like this one, however, you know better than to make anything of this result. You chalk it up to a lucky guess. Subsequently, you receive *another* newsletter from the

same investment-advice company telling you that IBM stock will go down the following month. When the stock does go down, you again chalk the prediction up to a lucky guess, but you do get a bit curious. When the third letter from the same company comes and predicts that IBM will go down again the next month, you do find yourself watching the financial pages a little more closely, and you confirm for the third time that the newsletter's prediction was correct. IBM has gone down this month. When the fourth newsletter arrives from the same company and tells you that the stock will rise the next month, and it actually does move in the predicted direction for the fourth time, it becomes difficult to escape the feeling that this newsletter is for real—difficult to escape the feeling that maybe you should send in the $29.95 for a year's worth of the newsletter. Difficult to escape the feeling, that is, unless you can imagine the cheap basement office in which someone is preparing next week's batch of 1,600 newsletters to be sent to 1,600 addresses: 800 of the newsletters predict that IBM will go up during the next month, and 800 of the newsletters predict that IBM will go down during the next month. When IBM does go up, that office sends out letters to *only* the 800 addressees who got the correct prediction the month before (400 predicting that the stock will go up in the next month and 400 predicting that it will go down, of course). Then you can imagine the "boiler room"—probably with telemarketing scams purring on the phones in the background—sending the third month's predictions to only the 400 who got the correct prediction the second week (200 predicting that the stock will go up in the next month and 200 predicting that it will go down). Yes, you were one of the lucky 100 who received four correct random predictions in a row! Many of these lucky 100 (and probably very impressed) individuals will pay the $29.95 to keep the newsletters coming.

Now this seems like a horrible scam to play on people. And indeed it is. But it is no less of a scam than when "respectable" financial magazines and TV shows present to you the "money manager who has beaten more than half his peers four years in a row!" Again, think back to our monkeys throwing the darts. Imagine that they were money managers making stock picks year after year. By definition, 50 percent of them will beat their peers during the first year. Half of these will again—by chance—beat their peers in the second year, making a total of 25 percent who beat their peers two years in a row. Half of these will again—by chance—beat their peers in the third year, making a total of 12.5 percent who beat their peers *three* years in a row. And finally, half of these 12.5 percent (i.e., 6.25 percent) will again beat their peers in the fourth year. Thus, about 6 of the 100 monkeys will have, as the financial shows and newspapers say, "consistently beaten other money managers for four years in a row!" These six monkeys who beat their dartboard peers (and, as we just saw, would beat a majority of *actual* Wall Street money managers; Ellis, 2016; Malkiel, 2016) certainly deserve spots on the financial television programs, don't you think?

Explaining Chance: Illusory Correlation and the Illusion of Control

[illegible]. [illegible] co-occurrences with great frequency, even when the two critical events are occurring [illegible]. In short, [illegible] (Kahneman, 2011; Whitson & Galinsky, 2008).

Controlled studies have demonstrated that [illegible]. Unfortunately, this finding generalizes to some real-world situations that adversely affect people's lives. For example, [illegible]. This is

the famous inkblot test in which the subject responds to blotches on a white paper. [illegible]. [illegible] blots. The problem with all of this is that there is no evidence that the Rorschach test [illegible] (Lilienfeld et al., 2010, 2012). [illegible]. [illegible].

Many of the interpersonal encounters in our lives have a large amount of chance in them: the blind date that leads to marriage, the canceled appointment that causes the loss of a job, the missed bus that leads to a meeting with an old high school friend. [illegible]. [illegible].

[illegible]. Psychologists have studied what has been termed the "illusion of control," that is, the tendency to believe that personal skill can affect outcomes determined by chance (Matute et al., 2011). Evidence of the widespread nature of this fallacy comes from the experience of states in which lotteries have been instituted. These states are descended on by purveyors of bogus books advising people how to "beat" the lottery—books that sell because people do not understand the implications of randomness. In fact, the explosion in the popularity of state lotteries in the United States did not occur until the mid-1970s, when New Jersey introduced participatory games in which players could scratch cards or pick their own numbers.

Chance and Psychology

In psychology, the tendency to try to explain everything, to have our theories account for every bit of variability rather than just the systematic nonchance components of behavior, accounts for the existence of many unfalsifiable psychological theories, both personal theories and those that are ostensibly scientific. [illegible]. [illegible]. The television talk-show guest (Chapter 4) who has an answer for every single case, for every bit of human behavior, should engender not admiration but suspicion. [illegible]

behavior, should engender not admiration but suspicion. [illegible].

Coincidence

The tendency to seek explanations for essentially chance occurrences leads to much misunderstanding regarding the nature of coincidental events. [illegible] bound to occur even if nothing other than chance is operating. Coincidences need no [illegible].

Most dictionary definitions of the word "coincidence" interpret it to refer to an accidental remarkable occurrence of related events. Because the same dictionaries define *accidental* as "occurring by chance," there is no problem here. [illegible]. Instead, they seek elaborate theories

in order to understand these events. [illegible] elaborate explanation of a coincidental event. On any given day, most of us probably [illegible]

If people truly understood what coincidence meant (a remarkable occurrence that is due to chance), they would not fall prey to the fallacy of trying to develop systematic, nonchance explanations for these chance events. Yet, [illegible]. For example, most of us have heard statements like "My goodness, what a coincidence! I wonder *why* that happened!" This reflects a fundamental error—coincidences do not [illegible]

Psychologist [illegible] (2001) [illegible]. One thing that contributes to the tendency to search for explanations of coincidental events is the mistaken idea that rare events never happen, that oddmatches are never due to chance. Our belief in this fallacy is intensified because probabilities are sometimes stated in terms of odds and because of the connotations that such statements have. Think of how we phrase the following: "Oh, goodness, that's very unlikely. The odds are 100 to 1 against that happening!" The manner in which we articulate such a statement strongly implies that it will *never* happen. Of course, we could say the same thing in a very different way, one that has very different connotations: "In 100 events of this type, this outcome *will* probably happen once." This alternative phrasing emphasizes that, although the event is rare, in the long run rare events *do* happen.

[illegible]. If you flipped 5 coins all at once and they all came up heads, you would probably consider this result an oddmatch, an unlikely event. You would be right. The probability of this happening in any one flip of 5 coins is 1/32 or 0.03. But if you flipped the 5 coins

100 times and asked how likely it is that in at least 1 of those 100 trials the coins would all come up heads, the answer would be 0.96. That is, in 100 trials, this rare event, this oddmatch, is *very likely* to happen.

In short, [illegible]. In August 1913, in a casino in Monte Carlo (Kaplan & Kaplan, 2007), black came up on a roulette wheel 26 times in a row! Or, take another example: If lotteries go on long enough, consecutive identical winning numbers are bound to be drawn eventually. For example, on June 21, 1995, in a German lottery called 6/49 (6 numbers are picked out of 49 possible) the numbers drawn were 15-25-27-30-42-48—exactly the same set of numbers that had been drawn on December 20, 1986 (Mlodinow, 2008). Many people were surprised to learn that over that time period the chance that *some* set of numbers would repeat was as high as 28 percent.

There are websites devoted to the "spooky" fact that many famous musicians died at age 27: Amy Winehouse, Kurt Cobain, Jim Morrison, Jimi Hendrix, Janis Joplin, and so on (O'Connor, 2011). Except that there is nothing "spooky" about it. It is not a fact in need of explanation. It is, instead, a random occurrence. The reason we know this is

because of a statistical analysis published in the *British Medical Journal* of 1,046 musicians who had a No. 1 album on the British charts from 1956 to 2007 (Barnett, 2011). The analysis indicated that there is no tendency for star musicians to die disproportionately at age 27.

[illegible]. Cognitive psychologist Daniel Kahneman (2011) describes how during the Yom Kippur War in 1973 he was approached by the Israeli Air Force for advice. Two squads of aircraft had gone out and one squad had lost four aircraft and the other had lost none. The Air Force wanted Kahneman to investigate whether there were factors specific to the different squadrons that were correlated with the outcome. But Kahneman knew that, with a sample this small, any such factors found would most likely be spurious—the result of mere chance fluctuation. Instead of doing a study, Kahneman used the insights in this chapter and told the Israeli Air Force not to waste their time. He says, "I reasoned that luck was the most likely answer, that a random search for a nonobvious cause was hopeless, and that in the meantime the pilots in the squadron that had sustained losses did not need the extra burden of being made to feel that they and their dead friends were at fault" (p. 116).

Personal Coincidences

Oddmatches that happen in our personal lives often have special meaning to us and, thus, we are especially prone not to attribute them to chance. There are many reasons for this tendency. Some are motivational and emotional, but others are due to failures of probabilistic reasoning. We often do not recognize that oddmatches are actually just a small part of a much larger pool of "nonoddmatches." It may seem to some of us that oddmatches occur with great frequency. But do they?

Consider what an analysis of the oddmatches in your personal life would reveal. Suppose on a given day you were involved in 100 distinct events. This does not seem an overestimate, considering the complexity of life in a modern industrial society. Indeed, it is probably a gross underestimate. You watch television, talk on the telephone, meet people, negotiate the route to work or to the store, do household chores, take in information while reading, send and receive emails and texts, complete complex tasks at work, and so on. All these events contain several components that are separately memorable. One hundred, then, is probably on the low side, but we will stick with it. An oddmatch is a remarkable conjunction of two events. How many possible different pairs of events are there in the 100 events of your typical day? Using a simple formula to obtain the number of combinations, we calculate that there are 4,950 different pair-

ings of events possible in your typical day. This is true 365 days a year.

Now, oddmatches are very memorable. You would probably remember for several years the day Uncle Bill called. Assume that you can remember all the oddmatches that happened to you in a ten-year period. Perhaps, then, you remember six or seven oddmatches (more or less, people differ in their criteria for oddness). What is the pool of nonoddmatches from which these six or seven oddmatches came? It is 4,950 pairs per day multiplied by 365 days per year multiplied by 10 years, or 18,067,500. In short, 6 oddmatches happened to you in 10 years, but 18,067,494 things that could have been oddmatches also happened. The probability of an oddmatch happening in your life is 0.00000033. It hardly seems strange that 6 out of 18 million conjunctions of events in your life should be odd. Odd things do happen. They are rare, but they do happen. Chance guarantees it (recall the example of simultaneously flipping five coins). In our example, six odd things happened to you. They were probably coincidences: remarkable occurrences of related events that were due to chance. Daniel Kahneman (2011) has argued that our language fails us here. We have terms for past thoughts that turned out to be true (premonition, intuition), but we have no words to mark and bring to our

attention past beliefs that turned out to be false. Most people would not spontaneously think to say, "I had a premonition that the marriage would not last, but I was wrong" (p. 202), because somehow that would seem strange to them. Without a word to mark the occurrence, we are not prone to mark all of our past predictions that failed to occur.

[illegible] oddmatches are commonly thought to be more odd than they really are. The famous "birthday problem" provides a good example of this. In a class of 23 people, what is the probability that 2 of them will have their birthday on the same day? What is the probability in a class of 35 people? Most people think that the odds are pretty low. Actually, in the class with 23 people, the odds are better than 50–50 that 2 people will have birthdays on the same day. And in the class of 35 students, the odds are very high (the probability is over 0.80). Thus, because there have been 45 presidents of the United States, it is not surprising that 2 (James Polk and Warren Harding) were born on the same day (November 2). Nor is it surprising, because 39 presidents have died, that 2 (Millard Fillmore and William Howard Taft) have died on the same day (March 8) and, furthermore, that three *more* (John Adams, Thomas Jefferson, James Monroe) have all also died on the same day (July 4!!).

Accepting Error in Order to Reduce Error: Clinical Versus Actuarial Prediction

The reluctance to acknowledge the role of chance when trying to explain outcomes in the world can actually decrease our ability to predict real-world events. [illegible]. But interestingly, [illegible]. It may seem paradoxical, but it is true that we must *accept* error in order to *reduce* error.

The concept that we must accept error in order to reduce error is illustrated by a very simple experimental task that has been studied for decades in cognitive psychology laboratories. The subject sits in front of two lights (one red and one blue) and is told that she or he is to predict which of the lights will be flashed on each trial and that there will be several dozen such trials (subjects are often paid money for correct

predictions). The experimenter has actually programmed the lights to flash randomly, with the provision that the red light will flash 70 percent of the time and the blue light 30 percent of the time. Subjects do quickly pick up the fact that the red light is flashing more, and they predict that it will flash on more trials than they predict that the blue light will flash. In fact, they predict that the red light will flash approximately 70 percent of the time. However, as discussed earlier in this chapter, subjects come to believe that there is a pattern in the light flashes and almost never think that the sequence is random. Instead, they switch back and forth from red to blue, predicting the red light roughly 70 percent of the time and the blue light roughly 30 percent of the time. Subjects rarely realize that—despite the fact that the blue light is coming on 30 percent of the time—if they stopped switching back and forth and predicted the red light every time, they would actually do better! How can this be?

Let's consider the logic of the situation. How many predictions will subjects get correct if they predict the red light roughly 70 percent of the time and the blue light roughly 30 percent of the time and the lights are really coming on randomly in a ratio of 70 to 30? We will do the calculation on 100 trials in the middle of the experiment—after

the subject has noticed that the red light comes on more often and is, thus, predicting the red light roughly 70 percent of the time. In 70 of the 100 trials, the red light will come on and the subject will be correct on about 70 percent of those 70 trials (because the subject predicts the red light 70 percent of the time). That is, in 49 of the 70 trials (70 times 0.70), the subject will correctly predict that the red light will come on. In 30 of the 100 trials, the blue light will come on, and the subject will be correct in 30 percent of those 30 trials (because the subject predicts the blue light 30 percent of the time). That is, in 9 of the 30 trials (30 times 0.30), the subject will correctly predict that the blue light will come on. Thus, in 100 trials, the subject is correct 58 percent of the time (49 correct predictions on red light trials and 9 correct predictions on blue light trials). But notice that this is a poorer performance than could be achieved if the subject simply noticed which light was coming on more often and then predicted it in every trial—in this case, noticing that the red light came on more often and predicting it in every trial (let's call this the 100 percent red strategy). Of the 100 trials, 70 would be red flashes, and the subject would have predicted all 70 of these correctly. Of the 30 blue flashes, the subject would have predicted none correctly but still would have a prediction accuracy of 70–12 percent better than the 58 percent correct that the subject achieved by switching back and forth.

The optimal strategy does have an implication though that troubles some people—that the optimal strategy will be wrong every time a blue occurs. And since blue light stimuli are occurring on at least *some* of the trials, to some people it just does not seem right *never* to predict them. [illegible]. In short, [illegible]. [illegible]; that is, [illegible].

Accepting error in order to make fewer errors is a difficult thing to do, however, as evidenced by the 60-year history of research on clinical versus actuarial prediction in psychology. [illegible] (i.e., aggregate predictions) that we discussed at the beginning of this chapter. [illegible]. More accurate predictions can be

made if we take more than one group characteristic into account (using the complex correlational techniques mentioned in Chapter 5—specifically a technique known as *multiple regression*). For example, predicting a life span of 60.2 years for people who smoke, are overweight, and do not exercise would be an example of an actuarial prediction based on a set of variables (smoking behavior, weight, and amount of exercise), and such predictions are almost always more accurate than predictions made from a single variable.

[illegible]. For example, in studies published in the *Journal of the American Medical Association* and in the *Annals of Internal Medicine* the following probabilistic trends were reported: people who are obese in middle age are four times more likely than nonobese people to have heart problems after age 65; overweight (but not obese) people are twice as likely to develop kidney problems; and obese people are seven times more likely to develop kidney problems (Seppa, 2006). But probabilistic prediction admits error. Not all obese people will have health problems. Recall the case (from Chapter 10) of the political broadcaster Tim

Russert who died of a heart attack at age 58. Physicians determined that Mr. Russert's probability of a heart attack in the next ten years was only 5 percent. That is, *most* people (95 out of 100) with Mr. Russert's profile would be heart-attack free for ten years. Mr. Russert was one of the unlucky 5 percent—he was an exception to the general trend.

People sometimes find it difficult to act on actuarial evidence, however, because doing so often takes mental discipline. For example, in 2003 the Food and Drug Administration issued a health-advisory warning of a potential link between a popular antidepressant drug and teen suicide. Many physicians worried that, on an actuarial basis, the warning would result in more suicides. The physicians acknowledged that fewer teenagers would die of suicide *because* of the drug, but they warned that *even more* children would die because of an increased hesitancy to prescribe the drug. This is indeed what happened. Treatment with this drug can put children at a temporary risk, but untreated depression is far worse. Most doctors thought that the warning would cost more lives than it would save (Dokoupil, 2007). That was the mathematics of the situation. Or perhaps we should say: That's the calculus of actuarial prediction. But it can be a hard calculus to follow when folk wisdom is saying things like "better to be safe than sorry." But in the domain of medical treatment "better to be safe than sorry" ignores one half of the equation. It focuses our attention on those who might be hurt by the treatment, but it totally ignores those who would be hurt if the treatment were unavailable.

[illegible]. In contrast, [illegible]. This is called *clinical,* or *case, prediction.* When engaged in clinical prediction, as opposed to actuarial prediction, professional psychologists claim to be able to make predictions about particular individuals that transcend predictions about "people in general" or about various categories of people. Clinical prediction would seem to be a very useful addition to actuarial prediction. There is just one problem, however. Clinical prediction doesn't work.

For clinical prediction to be useful, the clinician's experience with the client and her or his use of information about the client would have to result in better predictions than we can get from simply coding information about the client and submitting it to statistical procedures. In short, [illegible]

[redacted]. [redacted].

Research on the issue of clinical versus actuarial prediction has been consistent, and it has been going on for a long time. Since the publication in 1954 of Paul Meehl's classic book *Clinical Versus Statistical Prediction*, decades of research consisting of over a hundred research studies have shown that, in just about every clinical prediction domain that has ever been examined (psychotherapy outcome, parole behavior, college graduation rates, response to electroshock therapy, criminal recidivism, length of psychiatric hospitalization, and many more), [redacted] (Kahneman, 2011; Lewis, 2017; Morera & Dawes, 2006; Tetlock & Gardner, 2015). It is for this reason that some states in the United States have begun to replace subjective parole boards with actuarial methods when making prison release decisions (Walker, 2013).

In a variety of clinical domains, when a clinician is given information about a client and asked to predict the client's behavior, and when the same information is quantified and processed by a statistical equation that has been developed based on

actuarial relationships that research has uncovered, the actuarial prediction is more accurate than the clinician's prediction. In fact, even when the clinician has more information available than is used in the actuarial method, the latter is still superior. That is, when the clinician has information from personal contact and interviews with the client, in addition to the *same* information that goes into the actuarial equation, the clinical predictions still do not achieve an accuracy as great as the actuarial method.

Here we have an example of failing to "accept error in order to reduce error" that is directly analogous to the light prediction experiment previously described. Rather than relying on the actuarial information that the red light came on more often and predicting red each time (and getting 70 percent correct), the subjects tried to be correct on each trial by alternating red and blue predictions and ended up being 12 percent less accurate (they were correct on only 58 percent of the trials). Analogously, the clinicians in these studies believed that their experience gave them "clinical insight" and allowed them to make better predictions than those that can be made from quantified information in the client's file. In fact, their "insight" is nonexistent and leads them to make predictions that are worse than those they would make if they relied only on the public, actuarial information. It should be noted, though, that the superiority of actuarial prediction is not confined to psychology but extends to many other clinical sciences as well—for example, to medicine, and to fields such as financial services (Ellis, 2016; Lewis, 2017) and athletic coaching (Moskowitz & Wertheim, 2011).

[redacted] (pp. 373–374). [redacted]. For example, [redacted] (Dana et al., 2013). Instead, [redacted].

One anti-actuarial argument that is often raised is the old cliche that group statistics do not apply to single individuals, or to single events. But this is a vague and imprecise statement. Does the person making this argument think that if one is forced to play Russian roulette a single time and is allowed to select a gun with one or five bullets in the chamber, that you might as well pick the five rather than the one? It's a single, unique event, so it doesn't matter, right?

[illegible]. [illegible]. [illegible]. Although the field as a whole would have little to lose, individual practitioners who engage in

activities in the role of "experts" (i.e., in courtroom testimony) and imply that they have unique clinical knowledge of individual cases would, of course, lose prestige and perhaps income.

In fact, the field, and society, would benefit if we developed the habit of "accepting error in order to reduce error." In attempting to find unique explanations of every single unusual case (unique explanations that simply may not be possible given the present state of our knowledge), we often lose predictive accuracy in the more mundane cases. Recall the red–blue light experiment again. The "100 percent red strategy" makes incorrect predictions of all of the minority or unusual events (when the blue lights flash). What if we focused more on those minority events by adopting the "70-percent-red-30-percent-blue strategy"? We would now be able to predict 9 of those 30 unusual events (30 times 0.30). But the cost is that we lose our ability to predict 21 of the majority events. Instead of 70 correct predictions of red, we now have only 49 correct predictions (70 times 0.70). Predictions of behavior in the clinical domain have the same logic. In concocting complicated explanations for every case, we may indeed catch a few more unusual cases—but at the cost of losing predictive accuracy in the majority of cases, where simple actuarial prediction would work better.

Compulsive gamblers have a strong tendency not to "accept error in order to reduce error." For example, blackjack players had a tendency to reject a strategy called *basic* that is guaranteed to decrease the casino's advantage from 6 or 8 percent to less than 1 percent. Basic is a long-term statistical strategy, and the compulsive players tend to reject it because they believe that the best strategy should work every time and be keyed to the specifics of the situation. Instead of using an actuarial strategy that was guaranteed to save them thousands of dollars, compulsive gamblers are on a futile chase to find a way to make a clinical prediction based on the idiosyncrasies of each specific situation.

Another domain in which actuarial prediction often beats clinical prediction is sports. Many people saw the movie *Moneyball* in 2011, based on the book by Michael Lewis (2004). It told the story of Oakland A's manager Billy Beane, who overruled the "clinical" judgments of his baseball scouts (who tended to rely heavily on visible physical characteristics) and relied on statistics of past performance when evaluating potential team members. His teams overperformed relative to the money they spent, and the actuarial methods that he had borrowed from baseball statisticians were then copied by many other teams. Statistical methods have been shown to be superior to "coaches' judgments" in many other sports (see Moskowitz & Wertheim, 2011, for many examples).

Of course, [illegible].

[illegible]. Case information is highly useful in drawing attention to variables that are important and that need to be measured. [illegible]. First, [illegible]. Second, [illegible]—[illegible]. [illegible].

We will end this chapter, with the answer that psychologist Nicholas Epley (2013) gave to the interesting interview query: What's the question about your field that you dread being asked? Epley picked the classic question that psychologists often get asked in casual conversation, "Are you analyzing me?" The question reflects the Freud problem

discussed in Chapter 1. But Epley went on to explain that it was another aspect of the question that bothered him more; what he deemed a "deeper issue," and I agree. Epley claimed that the question "implies that I, as a psychologist, *could* indeed analyze you. [illegible]. In medicine, for instance, doctors prescribe drugs because the average outcome of those in the treatment group of a drug trial was better than the outcome of those in the placebo group…But as a psychologist, I often field questions that call on me to offer more individualized answers than our science can warrant."

Summary

The role of chance in psychology is often misunderstood by the lay public and by clinical practitioners alike. [illegible]. That is, [illegible]. Instead, [illegible]. [illegible], who sometimes mistakenly imply that clinical training confers an "intuitive" ability to predict an individual case. Instead, [illegible]. [illegible]. Thus, [illegible]. [illegible].

Chapter 12

The Rodney Dangerfield of the Sciences

Learning Objectives

12.1 Summarize the reasons why psychology suffers from a negative image

12.2 Explain why the interdisciplinary nature of psychology diminishes its scientific contributions

12.3 Outline the problems within the field of psychology that contributes to its negative image

12.4 Describe the issues that affect the field of psychology due to a growing ideological monoculture

12.5 Differentiate individual psychology from scientific psychology

12.6 Distinguish between scientific psychological research and pseudo-scientific claims

12.7 Summarize how psychology uses all of the components of science to understand the nature of human behavior

Rodney Dangerfield was a popular comedian for over three decades and whose trademark was the plaintive cry, “I don’t get no respect!” In a way, this is a fitting summary of psychology’s public image. This chapter will touch on some of the reasons that psychology appears to be the Rodney Dangerfield of the sciences.

Although there is a great public fascination with psychological topics, most

Although there is a great public fascination with psychological topics, most judgments about the field and its accomplishments are resoundingly negative. Psychologists are aware of this image problem, but most feel that there is little they can do about it, so they simply ignore it. This is a mistake. Ignoring psychology's image problem threatens to make it worse.

Psychology's Image Problem

Some of the reasons for psychology's image problem have already been discussed. For example, [illegible]. [illegible] (Overskeid, 2007). [illegible]. [illegible] (Gaynor, 2004), [illegible]

[illegible] (Freedman, 2012). However, [illegible].

Psychology and Parapsychology

The layperson's knowledge of reputable psychological research, outside of the work of Freud or Skinner, is virtually nonexistent. One way to confirm this fact is to look in your local bookstore to see what material on psychology is available to the general public. Inspection will reveal that the material generally falls into three categories. First, there will be a few classics (Freud, Skinner, Fromm, Erickson, etc.) heavily biased toward old-style psychoanalytic views that are totally unrepresentative of modern psychology. Frustratingly for psychologists, works of real worth in the field are often shelved in the science and/or biology sections of bookstores. For example, psychologist Steven Pinker's well-known and esteemed book *How the Mind Works* (1997) is often in the science section rather than the psychology section. Thus, the important work in cognitive science that he discusses becomes associated with biology, neurophysiology, or computer science rather than psychology. For example, in my local Barnes & Noble store, the Science section has subsections labeled Biology, Chemistry, Earth, and Physics, of course. But it also has a subsection labeled Cognitive Science, and in it are shelved some of the very best recent books on psychological research: Kahneman's (2011) *Thinking, Fast and Slow*; Wegner and Gray's (2016) *The Mind Club*; Gilovich and Ross's (2015) *The Wisest One in the Room*; and Tetlock and Gardner's (2015) *Superforecasting*. None of these books are shelved in the Psychology section at my Barnes & Noble and thus no one in the public will associate the first-rate psychological science in these volumes with the discipline of psychology itself.

The second class of material found in most stores might be called pseudoscience masquerading as psychology—that is, the seemingly never-ending list of so-called paranormal phenomena such as telepathy, clairvoyance, psychokinesis, precognition, reincarnation, biorhythms, astral projection, pyramid power, and psychic surgery. The presence of a great body of this material in the psychology sections of bookstores undoubtedly contributes to the widespread misconception that psychologists are the people who have confirmed the existence of such phenomena. There is a bitter irony for psychology in this misconception. In fact, the relationship between psychology and the paranormal is easily stated. These phenomena are simply not an area of active research interest in modern psychology. The reason, however, is a surprise to many people.

The statement that the study of ESP and other paranormal abilities is not accepted as part of the discipline of psychology will undoubtedly provoke the ire of many readers. Surveys have consistently shown that up to 40 percent of the general public

readers. Surveys have consistently shown that up to 40 percent of the general public believes in the existence of such phenomena and often holds these beliefs with considerable fervor (Poppy, 2017; Shermer, 2011). Like most religions, many of the so-called paranormal phenomena seem to promise things such as life after death, and for some people, they serve the same need for transcendence. It should not be surprising, then, that the bearer of the bad tidings that research in psychology does not validate ESP is usually not greeted with enthusiasm.

The statement that psychology does not consider ESP a viable research area invariably upsets believers and often provokes charges that psychologists are dogmatic in banishing certain topics from their discipline. But this criticism is wrong. Scientists do not determine by edict which topics to investigate. No proclamation goes out declaring what can and cannot be studied. Areas of investigation arise and are expanded or terminated according to a natural selection process that operates on ideas and methods. Those that lead to fruitful theories and empirical discoveries are taken up by a large number of scientists. Those that lead to theoretical dead ends or that do not yield replicable or interesting observations are dropped.

The reason that ESP, for example, is not considered a viable topic in contemporary psychology is simply that its investigation has not proved fruitful. Therefore,

very few psychologists are interested in it. It is important here to emphasize the word "contemporary," because the topic of ESP was of greater interest to psychologists some years ago, before the current bulk of negative evidence had accumulated. As history shows, research areas are not declared invalid by governing authorities; they are merely winnowed out in the competing environment of ideas.

ESP was never declared an invalid topic in psychology. The evidence of this fact is clear and publicly available (Galak et al., 2012; Hand, 2014; Nickell & McGaha, 2015). Many papers investigating ESP have appeared in legitimate psychological journals over the years. As recently as 2011, a major APA journal published a paper on a parapsychological effect (Bem, 2011). Alas, as is so often the case, the effects reported appear not to be reliable or replicable (Galak et al., 2012; Wagenmakers et al., 2011).

Parapsychologists who thrive on media exposure like to give the impression that the area is somehow new, thus implying that startling new discoveries are just around the corner. The truth is much less exciting. The study of ESP is actually as old as psychology itself. It is not a new area of investigation. It has been as well studied as many of the currently viable topics in the psychological literature. The results of the many studies that have appeared in legitimate psychological journals have been overwhelmingly negative. After more than 90 years of study, there still does not exist one example of an ESP phenomenon that is replicable under controlled conditions. In short, there is no demonstrated phenomenon that needs scientific explanation. For this reason alone, the topic is now of little interest to psychology.

Psychologists have played a prominent role in attempts to assess claims of paranormal abilities. Many of the most important books on the state of the evidence on paranormal abilities have been written by psychologists. But given that this is the case, it is ironic that psychology, the discipline that has probably contributed most to the negative assessment of ESP claims, is the field that is most closely associated with such pseudosciences in the public mind.

The Self-Help Literature

The third category in the bookstore psychology section is the so-called self-help literature. There are, of course, many different genres within this category (Lilienfeld, 2012; Meyers, 2008). Some books are spiritually uplifting tracts written with the purpose of generally increasing feelings of self-worth and competence. Others attempt to package familiar bromides about human behavior in new ways. A few (but all too few) are authored by responsible psychologists writing for the general public. Many that are

not in the latter category vie for uniqueness by presenting new "therapies" that are usually designed not only to correct specific behavioral problems but also to help satisfy general human wants (making more money, losing more weight, and having better sex are the "big three"), thereby ensuring larger book sales. These so-called new therapies are rarely based on any type of controlled experimental investigation. They usually rest on personal experience, or on a few case histories if the author is a clinician. This is often true of the treatments of so-called alternative medicine.

The many behavioral and cognitive therapies that have emerged after painstaking psychological investigation as having demonstrated effectiveness are usually poorly represented on the bookshelves. Lilienfeld (2012) estimates that of the 3,500 self-help books that are published each year, only about 5 percent of them have any scientific validation.

The situation is even worse in the electronic media and the internet. Radio and TV carry virtually no reports of legitimate psychology and instead present purveyors of bogus "therapies" and publicity-seeking media personalities who have no connection to the actual field of psychology. The main reason is that the legitimate psychological therapies do not claim to provide an instant cure or improvement, nor do they guarantee success or claim a vast generality for their effects ("Not only will you quit smoking, but every aspect of your life will improve!").

It is similar in the case of the internet. The lack of peer review ensures that the therapies and cures that one finds there are often bogus. Here is one example. In 2008, Paul Offit published an important book titled *Autism's False Prophets*, in which he detailed the many treatments for autism that have been found to be bogus by actual scientific research but that have enjoyed popularity among parents desperate for a treatment to help their children. One, facilitated communication, I have discussed in Chapter 6. Offit describes many other pseudoscientific treatments that have falsely raised parents' hopes and have led them to spend thousands of dollars and to waste their time and energy chasing a bogus "cure." On March 12, 2017, I identified one of the bogus chemical "cures" for autism discussed in Offit's book (I will not name it in order not to add to its publicity) and typed it and the word "autism" into Google. Of the first ten links that appeared in the outcome of my search, four links were to websites that were *advocating* this bogus chemical remedy.

Scientific accuracy is not guaranteed in a web search because websites are not peer reviewed. They thus provide no consumer protection for the random searcher with no further knowledge of the scientific literature on the topic in question. Advice given on television shows is little better (see Korownyk et al., 2014). Indeed, physicians are becoming more and more concerned about so-called cyberchondria (Peterson, 2012): people thinking that they are ill because of obsessive surfing on the web for negative symptoms. Indeed, the internet is full of bad medical advice—so much so that Google is working on search tools to remedy this problem (Mole, 2016). It is even more full of bad psychological advice.

[redacted]. As we showed in Chapter 4, [redacted]. [redacted]. As illustrated in Chapter 8, [redacted]. [redacted]

Recipe Knowledge

Finally, [redacted]. [redacted]

[illegible] Recipe knowledge is [illegible]. For example, [illegible]. This is recipe knowledge of the computer. Our knowledge of many technological products in our society is also recipe knowledge.

Of course, [illegible]. Indeed, [illegible]. In fact, the idea of recipe knowledge provides one way of conceptualizing the difference between basic and applied research. The basic researcher seeks to uncover the fundamental principles of nature without necessarily worrying about whether they can be turned into recipe knowledge. The applied researcher is more interested in translating basic principles into a product that requires only recipe knowledge.

[illegible] Now, [illegible]. However, a

[illegible]. Although a number of psychological researchers do work on turning basic behavioral principles into usable psychotherapeutic techniques, health-maintaining behavior programs, or models of efficient industrial organization, psychological research is largely basic research aimed at uncovering general facts and theories about behavior.

In all sciences, and in psychology in particular, there is usually a gap between the ideas that are productive for scientists and those that can be packaged to sell to the public. For example, there is legitimate research on "the power of positive thinking" in psychology (Sharot, 2011), but it bears little resemblance to the self-help prescriptions to that effect that were heard on *The Oprah Show*. Instead, the real psychological research literature is full of caveats, concerns about converging evidence, and the search for connectivity across research methods—in short, all of the real research concerns discussed in this book.

Consider the area of weight loss prescriptions. Scientists have slowly accumulated evidence for some mild prescriptions that help with weight control, but they are not breakthrough remedies. It is clear that the problem of obesity is complex and is subject to our warnings about multiple causation (Hewer, 2014; University of California, 2015b). The problem will clearly not have a single magic-bullet solution. Many scientists have stressed, for example, how the complexities of the food environment itself (advertising, portion sizes, marketing to children) contribute to the nation's obesity problem. The media want quick answers to questions that are of "public interest," whereas science produces slow answers to questions that are scientifically answerable—and all the questions that the public finds interesting might not be answerable.

Psychology and Other Disciplines

[illegible]. Many other allied disciplines, using a variety of different techniques and theoretical perspectives, also contribute to our knowledge. Many problems concerning behavior call for an interdisciplinary approach. However, a frustrating fact that most psychologists must live with is that when work on an interdisciplinary problem is publicized, the contributions of psychologists are often usurped by other fields.

[illegible]. For instance, the first major survey of the evidence on television's effects on children's behavior

the first major survey of the evidence on television's effects on children's behavior was conducted decades ago under the aegis of the U.S. Surgeon General. Thus, it is not surprising that the American Medical Association (AMA) passed a resolution to reaffirm the survey's findings of a suggested causal link and to bring the conclusions more publicity. Again, there is nothing wrong here, but an unintended consequence of the association of the findings on televised violence with the AMA is that it created the impression that the medical profession had conducted the scientific research summarized in the report. In fact, the overwhelming majority of the research studies on the effects of television violence on children's behavior have been conducted by psychologists. It was similar, decades later, when the American Academy of Pediatrics issued a report recommending limiting the internet use and cell phone use of children (Peterson, 2013). It was psychologists, not pediatricians, who did most of the scientific work that established these recommendations.

One of the reasons that the work of psychologists is often ascribed to other disciplines is that the word *psychologist* has, over the years, become ambiguous. Many research psychologists commonly append their research specialty to the word *psychologist* when labeling themselves, calling themselves, for example, physiological psychologists, cognitive psychologists, industrial psychologists, evolutionary

psychologists, or neuropsychologists. Some use a label that does not contain a derivative of the word "psychology" at all, for example, neuroscientist, cognitive scientist, artificial intelligence specialist, and ethologist. Both of these practices—in conjunction with the media's bias that "psychology isn't a science"—lead to the misattribution of the accomplishments of psychologists: The work of physiological psychologists is attributed to biology, the work of cognitive psychologists is attributed to computer science and neuroscience, the work of industrial psychologists is attributed to engineering and business, and so on. [illegible] Of course, [illegible]. Author Michael Lewis (2017), who wrote a book on Kahneman's work, admits that it is quite natural for a layperson to ask [illegible].

Finally, psychology departments are often misunderstood even within their own universities. Susan Putnam, chair of the psychology program at Canisius College, described how she battled to get psychology classified as a science at her institution (Weir, 2015). [illegible]

Our Own Worst Enemies

Lest it appear that we are blaming everyone else for psychology's image problems, it is about time that we acknowledge the contribution of psychologists themselves to confusion about their field. [illegible]. However, the focus of this section is on a different problem altogether: [illegible].

[illegible]. This attitude presents a serious threat to the integrity of psychotherapy. First, there is the proliferation of therapies that has occurred because of a reluctance to winnow out those that do not work. Such a proliferation not only removes a critical consumer protection but also promotes confusion in

the field. Second, there is an inconsistency in a therapeutic community that, on the one hand, argues against scientific evaluation because it is "more art than science," in the common phrase, but is still greatly concerned about what is the 800-pound gorilla in the room: reimbursement for services by government and private health insurers. [redacted].

Some readers of the first few editions of this book commented that they thought I had "let psychologists get off too easily" by not emphasizing more strongly that unprofessional behavior and antiscientific attitudes among psychologists themselves contribute greatly to the discipline's image problem. In trying to provide more balance here, I have relied heavily upon the work of Robyn Dawes (1994) and Scott Lilienfeld (2012; Lilienfeld et al., 2014). Dawes does not hesitate to air psychology's dirty linen and, at the same time, to argue that the scientific attitude toward human problems that is at the heart of the true discipline of psychology is of great utility to society. For example, [redacted] (1994, p. vii). Likewise, Lilienfeld (2012) argues

that [redacted] (pp. 122–123).

What Dawes and Lilienfeld are objecting to is that [redacted]. For example, [redacted] (see Chapters 10 and 11). [redacted] surrounding clinical psychology [redacted]. When pushed to defend licensure requirements as anything more than restraint of trade, however, the organization uses its scientific credentials as a weapon (one president of the APA, defending the organization from attack, said "Our scientific base is what sets us apart from the social workers, the counselors, and the Gypsies"; Dawes, 1994, p. 21). But the very methods that the field holds up to justify its scientific status have revealed that the implication that licensed psychologists have a unique "clinical insight" is false (Tracey et al., 2014). [redacted].

Several categories of pseudoscience have flourished in clinical psychology during the past few decades, including: unvalidated and bizarre treatments for trauma; demonstrably ineffective treatments for autism such as facilitated communication (see Chapter 6); the continued use of inadequately validated assessment instruments (e.g., many projective tests); and the use of highly suggestive therapeutic techniques to unearth memories of child abuse (Baker et al., 2009; Lilienfeld, 2007, 2013).

[redacted] (Lilienfeld et al., 2014). The list is so long that we can only give a small sampling here. For example, the critical-incident stress debriefing has, in many localities, become a standard procedure used to treat witnesses to catastrophic and traumatic events such as bombings, shootings, combat, terrorism, and earthquakes (Foa et al., 2013; McNally et al., 2003). The debriefing procedure involves having the client "talk about the event and ventilate their emotions, especially in the company

of peers who have experienced the same incident" (McNally et al., 2003, p. 56), and its purpose is to reduce the incidence of posttraumatic stress disorders (PTSDs). The majority of debriefed clients report that the experience was helpful. Of course, no one who has read this book will find that evidence convincing (recall the discussion of placebo effects in Chapter 4). A control group (which is not given the critical-incident stress debriefing) is obviously needed. In fact, "the vast majority of trauma survivors recover from initial posttrauma reactions without professional help" (McNally et al., 2003, p. 45), so it clearly needs to be demonstrated that the recovery rate is higher when the critical-incident stress debriefing is used. Properly controlled studies have shown that this is not the case (Foa et al., 2013; McNally et al., 2003), yet the procedure continues to be used.

Emery et al. (2005), in a review of a large body of evidence, have shown that, likewise, [redacted] (Novotney, 2008). For example, they describe several assessment instruments used by clinical psychologists purportedly to assess children's best interests in these custody disputes. After reviewing several of these instruments—for example, scales purporting to assess the perception of relationships and parental awareness

skills—Emery et al. (2005) conclude that none of them have demonstrated reliability or validity. They note that "no study examining the properties of these measures has ever been published in a peer-reviewed journal—an essential criterion for science" (p. 8) and conclude that "our bottom-line evaluation of these measures is a harsh one: these measures assess ill-defined constructs, and they do so poorly, leaving no scientific justification for their use in child custody evaluations" (p. 7).

Things may be looking up, however. [redacted] (Baker et al., 2009, p. 67). This report received considerable publicity, but some of the discussion in the general media confused the issue as much as clarified it. An otherwise accurate report in *Newsweek* magazine was unfortunately titled "Ignoring the Evidence: Why Do Psychologists Reject Science?" (Begley, 2009). The title mistakenly implies that it is *all* psychology that rejects science, rather than the problematic subfield of clinical psychology. This confusing title is bitterly ironic given that the logic of the APS report was that of all the rest of psychology—which *does* adhere to the scientific method—speaking in distress to just *one* of its many subfields that does not (clinical psychology).

[redacted]. This Jekyll and Hyde aspect of the discipline was clearly apparent in the recovered-memory–false-memory debate of the last two decades (Lilienfeld, 2007; Loftus & Guyer, 2002; McHugh, 2008; Patihis et al., 2014). Many cases were reported of individuals who had claimed to remember instances of child abuse that had taken place decades earlier but had been forgotten. Many of these memories occurred in the context of therapeutic interventions. It is now clear that some of these memories were induced by the therapy itself (Ammirati & Lilienfeld, 2015; Lilienfeld, 2007; Loftus & Guyer, 2002). In the emotionally charged atmosphere of such an explosive social issue, psychologists provided some of the more balanced commentary and, most important,

psychologists provided some of the more balanced commentary and, most important, some of the more dispassionate empirical evidence on the issue of recovered or false memories (Brainerd & Reyna, 2005; McNally & Geraerts, 2009; Moore & Zoellner, 2007; Patihis et al., 2014).

Here we have the Jekyll and Hyde feature of psychology in full-blown form. Some of the cases of therapeutically induced false memories—and, hence, of the controversial phenomenon itself—were caused by incompetent and scientifically ignorant therapists who were psychologists. On the other hand, the resolution of the controversy we do have is in large part due to the painstaking efforts of research psychologists who studied the relevant phenomena empirically. Finally, [illegible] (Kenney, 2008; Novella, 2015).

In his book on research methods, psychologist Douglas Mook (2001) referred to my use of the Rodney Dangerfield joke to title of this chapter and commented that "[illegible]" (p. 473). I agree completely with this sentiment. Mook is right that the student of psychology needs to understand the paradoxes that

surround the discipline. As I have presented it in this book, as the science of human behavior, the discipline of psychology often gets too little respect. But the face that psychology often presents to the public—that of a clinician claiming "unique" insight into people that is not grounded in research evidence—often gets too much respect. The discipline is often represented to the public by segments of psychology that do not respect its unique defining feature—that it validates statements about human behavior by employing the methods of science.

There is another way though, that psychology might be said to be getting too much respect, and I will deal with this in the next section. It is an aspect of modern psychology that threatens the objectivity of the discipline.

Our Own Worst Enemies, Part II: Psychology Has Become an Ideological Monoculture

As I have mentioned, the discussion in the previous section was motivated by feedback I have received from some readers who thought that this book was too positive about psychology. These readers of earlier editions of the book thought that I had "let psychology off the hook," so to speak, because I did not say enough about the flaws within the discipline. The primary thing that these readers pointed me to were the anti-scientific attitudes within psychology itself—primarily within clinical psychology. The previous section, which has appeared in several recent editions, was my attempt to accommodate the feedback of these critics of earlier editions.

[illegible] Indeed, in earlier chapters I have pointed out some flaws in psychology as a science that are generic to the discipline in that they span many of its subspecialty areas. For example, in the very first chapter I admitted that [illegible]. Then, in a later chapter, I also pointed out that [illegible] (Gilbert et al., 2016; Open Science Collaboration, 2015). [illegible] I also cited a study indicat-

[illegible]. I also cited a study indicating that [illegible] (Fanelli, 2010). All of these are flaws that indicate that [illegible].

Many of these criticisms have been problems within psychology for some time. However, the issue I will discuss in the remainder of this section is a problem that has intensified in the last couple of decades. It is a problem that is much more seriously impeding psychology in 2018 than it was in 1986, when the first edition of this book appeared. That problem is the ideological homogenization of psychology as a discipline.

It has always been true that [illegible]. Even 30 or 40 years ago, there were more liberal psychology professors than there were conservative ones—more Democrats than there were Republicans. But much converging research has shown that this imbalance has become even more marked in the last 20 years (Duarte et al., 2015)—so much so, that it would not be unfair to characterize the field of psychology as an ideological monoculture. Studies of social science departments in universities have indicated that 58–66% of professors identify themselves as liberals and just 5–8% as conservatives (Duarte et al., 2015).

The imbalance in psychology departments is even worse, with 84% of professors identifying as liberal and just 8% as conservative. This imbalance has skyrocketed in recent years. In 1990, the ratio was 4 liberals for every 1 conservative in psychology departments (the ratio is 1 liberal to 2 conservatives in the entire US population)—a strong imbalance, but still, the 20% of faculty who were conservatives at least provided some diversity. But by the year 2000, the ratio had climbed to 6 liberals for every 1 conservative (Duarte et al., 2015). And by the year 2012, the ratio had risen to an astonishing 14 to 1—virtually an ideological monoculture.

It is true that an ideological imbalance will not be a problem for many areas of psychology. [illegible]. So we are not suggesting here that all areas of research in psychology have this problem, or even a majority of them. Nevertheless, [illegible]. For example, [illegible].

It should be clear why an ideological imbalance is a problem in areas of research like those I have just enumerated. In Chapter 2, I discussed the unique feature of science that allows it to overcome the myside bias of individual scientists. Recall from that discussion that I emphasized that [illegible] (that they are completely objective or that they are never biased), [illegible].

However, [illegible] tion and cross-checking. What ruins it is when all of the investigators share exactly the [illegible] Then, the social process of error checking and correction cannot work as I described it in Chapter 2. Unfortunately, [illegible]. The pool of investigators is politically

homogeneous, and thus we cannot rest assured that there is enough variability in our science to objectively approach politically charged topics like those mentioned above.

It would be a mistake for psychologists to think that there are easy ways around this homogeneity—for instance, that they could just try harder to be objective on an individual basis. This would amount to denying what I stated in Chapter 2: that scientists are not uniquely virtuous in their objectivity; instead, they are kept honest by the *social process* of science. An ideological monoculture will not keep psychology honest in this way, because it removes the social milieu of criticism and cross-checking. Ironically enough, there is a well-known psychological phenomenon that suggests that it would be all too tempting for psychologists to think that they do not have this problem of myside bias—that they can set aside their ideological preferences while doing their science. The phenomenon is called the bias blind spot, which is the label for the finding that it is relatively easy for people to recognize bias in the decisions of others, but it is difficult to detect bias in their own judgments (Pronin, 2007). It would be all too easy for psychologists to think (wrongly) that they are immune from the bias blind spot and that ideological homogeneity is not a problem for their field.

There is an additional reason why it would be tempting for psychologists to wrongly assume that they have a unique ability to avoid bias. As the statistics presented

above show, the vast majority of psychologists are liberal Democrats. Liberal research psychologists have become accustomed, as we all have, to media presentations that are critical of conservative Republicans who do not accept the conclusions of climate science, or of evolutionary biology. These media presentations are correct, of course. The role of human activity in climate change is established science, and evolution is a biological fact. Thus, the denial of climate science or of evolutionary science clearly has a negative connotation, and rightly so. However, there is a trap lying in wait for liberal psychologists here. It is a very tempting step to say to oneself: Well, I get climate science right and Republicans get it wrong; and I get evolution right and conservative Republicans get it wrong; so therefore we liberal psychologists are getting everything right about psychology too (again, think of all the charged topics we mentioned above: parenting, sexuality, crime, poverty, etc.).

In short, psychologists might say to themselves: "Well, we may all be Democrats with no political variability among us, but that doesn't matter because the Republicans deny science and we are the party of science." That is pretty much what the Democratic Party did years ago when it declared itself the "party of science" and labeled the Republican Party as the science deniers. That stance spawned a series of books with titles like *The Republican War on Science* (Mooney, 2005). This might have been a fine political strategy for the Democratic Party, but research psychologists should know better. They should be able to see the obvious selection effects operating here—namely, that the issues in question (climate science and creationism/evolution) are cherry-picked for reasons of politics and media interest. In order to *correctly* call one party the party of science and the other the party of science deniers, one would of course have to have a representative sampling of scientific issues to see whether members of one party are more likely to accept the scientific consensus.

In fact, [illegible]. [illegible]. In fact, and ironically, there are enough examples to produce a book parallel to the Mooney volume cited above titled *Science Left Behind: Feel-Good Fallacies and the Rise of the Anti-Scientific Left* (Berezow & Campbell, 2012). We mentioned two of these in earlier chapters: [illegible] consensus in psychological science that intelligence is moderately heritable (Deary, 2013; Plomin et al., 2016); [illegible] (Bertrand

et al., 2010; Black et al., 2008; CONSAD, 2009; Kolesnikova & Liu, 2011; O'Neill & O'Neill, 2012; Solberg & Laughlin, 1995).

These aren't the only two issues, though. [redacted] (just as conservatives obfuscate the research on global warming) [redacted] (Chetty et al., 2014; McLanahan et al., 2013; Murray, 2012); [redacted] (Seidenberg, 2017); [redacted] 2015); [redacted] (National Academies of Sciences, 2016); [redacted] (Vogel, 2016); [redacted] (Pinker, 2002); [redacted].

I will stop here because the point is made. There is plenty of science denial on the liberal side to balance the anti-scientific attitudes of conservatives toward climate change and evolutionary theory. [redacted]. [redacted]. The litany of examples I have laid out here should be enough to sober psychologists and stop them from thinking that they are not subject to the bias blind spot that they themselves have discovered (Pronin, 2007).

Duarte et al. (2015) provide several examples indicating that [redacted]. They discuss a study that attempted to link the conservative worldview with "the denial of environmental realities." Subjects were presented with the following item: If things continue on their present course, we will soon experience a major environmental catastrophe. If the subject did not agree with this statement, they were scored as denying environmental realities. But as Duarte et al. point out, the term "denial" implies that what is being denied is a descriptive fact. However, without a clear description of what "soon" means in this statement, or "major" means, or what "catastrophe" means, the statement itself is not a fact—and so labeling one set of respondents as deniers reflects little more than the ideological biases of the study's authors. Other statements in the questionnaire have a similar logic. If the subjects did not agree with vague environmental values and instead affirmed statements like "the balance of nature is strong enough to cope with the impacts of modern industrial nations" then they were coded as denying environmental "realities."

Another study discussed by Duarte et al. tried to link aspects of the conservative temperament to making unethical decisions. One item was a very short and vague scenario involving one employee sending a sexist email to another employee (Felicity) after the two had had a work disagreement. The subject was asked to take the position of a manager not involved in the incident and decide whether the manager should write a letter supporting Felicity in her sexual harassment complaint. The authors of the study coded as ethical behavior the manager immediately writing the letter. Anything less than that choice on the part of the subject was coded as less ethical, or even unethical. Because so little information was given about the case, this item measured little more than the subject's pre-existing bias toward one party or the other in a sexual harassment case, yet the study was billed as an examination of *unethical* decision-making. Just like the environmental study in the previous paragraph, this

research was displaying the tendency to mark legitimate policy differences as absolutely right (or ethical) or absolutely wrong (or unethical)—with a bias toward identifying the "right" response as the liberal one!

This tendency to conflate liberal responses with the right response (or ethical response, or fair response, or open-minded response) is particularly prevalent in the subareas of social psychology and personality psychology. It often takes the form of labeling any legitimate policy difference with liberalism as some kind of intellectual or personality defect (dogmatism or authoritarianism or racism or prejudice). This has been true for many years in the study of racism by social psychologists. Many of the different scales in use include items on policy issues such as affirmative action, crime prevention, busing to achieve school integration, or attitudes toward welfare reform. Someone with legitimate policy differences with affirmative action or busing—or someone indicating that they are concerned about crime—will almost always be scored in the racist direction on these scales (Snyderman & Tetlock, 1986; Tetlock, 1994). In such studies, the overt ideological bias of psychology is blatantly obvious to any neutral observer. The purpose of such studies seems, transparently, to be to label anyone that does not adhere to liberal orthodoxy as a racist.

In fact, there is a whole subspecialty area in social psychology devoted to showing that negative traits such as prejudice and stereotyping and unfairness are

associated with the conservative temperament. There is even a theory—the "intrinsic thesis"—that hypothesizes that the increasing political polarization surrounding scientific issues is due to the "psychological deficiencies among conservatives as compared to liberals" (p. 36, Nisbet et al., 2015). Recently, there have been a flurry of psychological studies purporting to prove that conservatives are more prejudiced, less open-minded, and indeed less intelligent than liberals. The problem is that many of these studies have not been replicable, were poorly designed, or were designed and interpreted in a biased manner (Brandt et al., 2014; Chambers et al., 2013; Crawford, 2012; Duarte et al., 2015; Jussim et al., 2016; Kahan, 2013; Nisbet et al., 2015; Oswald et al., 2013).

And then there are the legion of studies that are overly hyped or misleadingly hyped because they seem to support a liberal conclusion. The classic example is the work on stereotype threat (see Jussim et al., 2016), which has been reported incorrectly in many media outlets and psychology textbooks. The actual finding was that introducing a stereotype threat increased the test score difference between African-American and white college students (Jussim et al., 2016). Because the original authors used a confusing statistical reporting procedure, textbooks often report (incorrectly) that the study found that racial group test score differences are eliminated when stereotype threat is removed. That is not at all the finding, but it is the one that was propagated widely because of psychology's ideological monoculture.

The examples could continue (see Duarte et al., 2015; Jussim et al., 2016), but I will stop here to stress that all of this does psychology no good. This ideological bias in the discipline is becoming more discernible to the general public. Indeed the psychology monoculture has made true the old joke: Psychology departments exist so that Democrats can say "studies show." More seriously for the discipline, it will certainly be true that funding agencies will become more aware of the ideological bias, as will the state legislatures funding the individual departments at their state universities. None of this will be good for psychology.

Our ideological monoculture has prevented us from challenging the host of quasi-psychological concepts that have been used in the last decade to suppress free speech on university campuses (Lukianoff & Haidt, 2015). Most of these concepts—concepts such as safe spaces, trigger warnings, rape culture, or micro-aggressions—have no empirical or theoretical grounding in psychological research. Yet psychologists have not been prominent in explaining to university students and administrators that these concepts are not grounded in psychological science. Indeed, many psychologists have

been high-level administrators in some of the universities that are proliferating these concepts like mental viruses. A notable exception is Scott Lilienfeld's (2017) thorough explication of what it would take, research-wise, to properly ground the concept of micro-aggression—to change it from its present status as a mere political weapon into a behavioral science concept. Not surprisingly for those of you who have read Chapter 6 (Clever Hans, etc.), he recommends changing the term to something more neutral that does not carry so much theory with it.

Finally, psychology's image is not helped by the tendency of one of its organizations, the American Psychological Association, to go way beyond science into politics and social advocacy with its public policy statements (the other major psychological association—the Association for Psychological Science—largely avoids this tendency). In a perceptive article on psychology's image problem, Ferguson (2015) discusses how the APA's public policy statements have repeatedly strayed into politics and have reinforced the public view that the organization is not scientific but an advocacy organization (politically, for liberal and Democratic positions). Ferguson discusses APA policy statements on abortion and welfare reform as particularly problematic—more politics than science. The ideological monoculture in university psychology departments is thus mirrored by the most publicly visible organization representing psychologists, the APA.

Isn't Everyone a Psychologist? Implicit Theories of Behavior

We all have theories about human behavior. It is hard to see how we could get through life if we did not. In this sense, we are all psychologists. It is very important, though, to distinguish between this individual psychology and the type of knowledge produced by the science of psychology. The distinction is critical because the two are often deliberately confused in popular writings about psychology, as we shall see.

Much of our personal psychological knowledge is recipe knowledge. We do certain things because we think they will lead others to behave in a certain way. We behave in particular ways because we think that certain behavior will help us achieve our goals. But it is not the mere presence of recipe knowledge that distinguishes personal psychology from scientific psychology (which also contains recipe knowledge). The main difference here is that the science of psychology seeks to validate its recipe knowledge. Scientific evaluation is systematic and controlled in ways that individual validation procedures can never be.

In addition, science always aspires to be more than recipe knowledge of the natural world. Scientists seek more general, underlying principles that explain why the recipes work. Rather than being coherently constructed, many people's personal psychological theories are merely a mixture of platitudes and clichés, often mutually contradictory, that are used on the appropriate occasion. They reassure people that an explanation does exist and, furthermore, that the danger of a seriously contradictory event—one that would deeply shake the foundations of a person's beliefs—is unlikely to occur. As discussed in Chapter 2, although these theories may indeed be comforting, comfort is all that theories constructed in this way provide. In explaining everything post hoc, these theories predict nothing. By making no predictions, they tell us nothing. Theories in the discipline of psychology must meet the falsifiability criterion, and in doing so, they depart from the personal psychological theories of many laypeople. Theories in psychology can be proved wrong, and, therefore, they contain a mechanism for growth and advancement that is missing from many personal theories.

The Source of Resistance to Scientific

Psychology

For the reasons we just discussed, it is important not to confuse the idea of a personal psychological theory with the knowledge generated by the science of psychology. Such confusion is often deliberately fostered to undermine the status of psychology in the public mind. The idea that "everyone's a psychologist" is true if it is understood to mean simply that we all have implicit psychological theories. But it is often subtly distorted to imply that psychology is not a science.

We discussed in Chapter 1 why the idea of a scientific psychology is threatening to some people. A maturing science of behavior will change the kinds of individuals, groups, and organizations that serve as sources of psychological information. It is natural that individuals who have long served as commentators on human psychology and behavior will resist any threatened reduction in their authoritative role. Chapter 1 described how the advance of science has continually usurped the authority of other groups to make claims about the nature of the world. The movement of the planets, the nature of matter, and the causes of disease were all once the provinces of theologians, philosophers, and generalist writers. Astronomy, physics, medicine, genetics, and other sciences have gradually wrested these topics away and placed them squarely within the domain of the scientific specialist.

The issue, then, is the changing criteria of belief evaluation. Few newspaper editorials ever come out with strong stands on the composition of the rings of Saturn. Why? No censor would prevent such an editorial. Clearly the reason it is not written is that it would be futile. Society knows that scientists, not editorial writers, determine such things. Some people find it difficult to accept such a state of affairs when it comes to psychology. They cling tenaciously to their right to declare their own opinions about human behavior even when these opinions contradict the facts. Of course, the correct term here is really not "right," because, obviously, in a free society, everyone has the *right* to voice opinions, regardless of their accuracy. It is important to understand that what many people want is much *more* than simply the right to declare their opinions about human behavior. What they really want is *the conditions that are necessary for what they say to be believed.* When they make a statement about human psychology, they want the environment to be conducive to the acceptance of their beliefs. This is the reason that there are always proponents of the "anything-goes" view of psychology; that is, the idea that psychological claims cannot be decided by empirical means and are simply a matter of opinion. But science is always a threat to the "anything-goes" view, because it has a set of strict requirements for determining whether a knowledge claim is to be believed. Anything does not go in science. This ability to rule out false theories and facts accounts for scientific progress.

In short, a lot of the resistance to scientific psychology is due to what might be termed "conflict of interest." As discussed in earlier chapters, many pseudosciences are multimillion-dollar industries that thrive on the fact that the public is unaware that statements about behavior can be empirically tested. The public is also unaware that many of the claims that are the basis of these industries (such as astrological prediction, subliminal weight loss, biorhythms, facilitated communication, and psychic surgery) have been tested and found to be false. Unproven medical remedies end up costing the public more than is spent on legitimate medical research (Mielczarek & Engler, 2012).

How do we recognize pseudoscientific claims? Clinical psychologist Scott Lilienfeld (2005, p. 40) gives us a list of things to watch for that could serve as a summary of many of the things that have been covered in this book. Pseudoscientific claims tend to be characterized by:

- A tendency to invoke ad hoc hypotheses as a means of immunizing claims from falsification
- An emphasis on confirmation rather than refutation
- A tendency to place the burden of proof on skeptics, not proponents, of claims

- Excessive reliance on anecdotal and testimonial evidence to substantiate claims
- Evasion of the scrutiny afforded by peer review
- Failure to build on existing scientific knowledge (lack of connectivity)

True scientists are at pains to emphasize these criteria rather than to avoid them. In response, the pseudoscience industry continues to oppose the authority of scientific psychology to evaluate behavioral claims. However, the purveyors of pseudoscience often do not need to do direct battle with psychology. They simply do an end run around psychology and go straight to the media with their claims. The media make it very easy for cranks, quacks, and pseudoscientists to get publicity for their claims without going through the peer review process. The talk shows that have inundated the airwaves do not ask the guests to produce their bibliographies of scientific research. If these guests are "interesting," they are simply put on the show. And the internet is no better. Anyone can put up a website claiming—and selling—anything. Websites are not peer reviewed, to say the least!

Science, then, does rule out knowledge claims that do not meet the necessary tests. The courts rule out claims of knowledge too. In ruling on a famous case known as *Daubert vs. Merrell Dow*, the Supreme Court established when expert testimony could be presented in court—that is, what makes expert testimony *expert*! The Court identified four factors that judges should consider when deliberating about whether

to allow expert testimony: (a) The "testability" of the theoretical basis for the opinion; (b) The error rates associated with the approach, if known; (c) Whether the technique or approach on which the opinion is based has been subjected to peer review; and (d) Whether the technique or approach is generally accepted in the relevant scientific community (Emery et al., 2005; Michaels, 2008). The four criteria map onto major topics in this book: (a) falsifiability; (b) probabilistic prediction; (c) public knowledge subjected to peer review; and (d) scientific knowledge based on convergence and consensus. The courts, like science, have ruled out claims of special knowledge, intuition, and testimonials as adequate evidence.

In this book, we have briefly touched on what are considered adequate and inadequate tests in science. Introspection, personal experience, and testimonials are all considered inadequate tests of claims about the nature of human behavior. Thus, it should not be surprising that conflict arises because these are precisely the types of evidence that nonpsychologist commentators have been using to support their statements about human behavior since long before a discipline of psychology existed.

However, it should not be thought that I am recommending a sour, spoilsport role for the science of psychology. Quite the contrary. The actual findings of legitimate psychology are vastly more interesting and exciting than the repetitious gee-whiz pseudoscience of the media. Furthermore, it should not be thought that scientists are against fantasy and imagination. However, we want fancy and fantasy when we go to the movies or the theater—not when we go to the doctor's office, buy insurance, register our children for child care, fly in an airplane, or have our car serviced. We could add to this list: going to a psychotherapist, having our learning-disabled child tested by a school psychologist, or taking a friend to suicide-prevention counseling at the university psychology clinic. Psychology, like other sciences, must remove fantasy, unfounded opinion, "common sense," commercial advertising claims, the advice of gurus, testimonials, and wishful thinking from its search for the truth.

It is difficult for a science to have to tell parts of society that their thoughts and opinions are needed—but not here. Psychology is the latest of the sciences to be in this delicate position. The difference in time period for psychology, however, is relevant. Most sciences came of age during periods of elite control of the structures of society, when the opinion of the ordinary person made no difference. Psychology, on the other hand, is emerging in a media age of democracy and ignores public opinion at its own peril. Many psychologists are now taking greater pains to remedy the discipline's lamentable record in public communication. As more psychologists take on a public

lamentable record in public communication. As more psychologists take on a public communication role, the conflicts with those who confuse a personal psychology with scientific psychology are bound to increase.

Not everyone is a physicist, even though we all hold intuitive physical theories. But in giving up the claim that our personal physical theories must usurp scientific physics, we make way for a true science of the physical universe whose theories, because science is public, will be available to us all. Likewise, everyone is not a psychologist. But the facts and theories uncovered by the science of psychology are available to be put to practical ends and to enrich the understanding of all of us.

The Final Word

We are now at the end of our sketch of how to think straight about psychology. It is a rough sketch, but it can be of considerable help in comprehending how the discipline of psychology works and in evaluating new psychological claims. Our sketch has revealed the following:

1. Psychology progresses by investigating solvable empirical problems. This progress is uneven because psychology is composed of many different subareas, and the problems in some areas are more difficult than in others.

2. Psychologists propose falsifiable theories to explain the findings that they uncover.
3. The concepts in the theories are operationally defined, and these definitions evolve as evidence accumulates.
4. These theories are tested by means of systematic empiricism, and the data obtained are in the public domain, in the sense that they are presented in a manner that allows replication and criticism by other scientists.
5. The data and theories of psychologists are in the public domain only after publication in peer-reviewed scientific journals.
6. What makes empiricism systematic is that it strives for the logic of control and manipulation that characterizes a true experiment.
7. Psychologists use many different methods to arrive at their conclusions, and the strengths and weaknesses of these methods vary.
8. The behavioral principles that are eventually uncovered are almost always probabilistic relationships.
9. Most often, knowledge is acquired only after a slow accumulation of data from many experiments each containing flaws but nonetheless converging on a common conclusion.

The most exciting endeavor in science today is the quest to understand the nature of human behavior. By learning the concepts in this book you become able to follow this quest and perhaps, indeed, become a part of it!

References

Adler, J. (2006, November 6). Plotting Pluto's comeback. *Newsweek*, pp. 60–61.

AEI/Brookings Working Group on Poverty (2015). *Opportunity, responsibility, and security: A consensus plan for reducing poverty and restoring the American Dream*. The American Enterprise Institute for Public Policy Research (AEI) and the Brookings Institution.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.; Text Revision). Washington, DC: Author.

Ammirati, R., & Lilienfeld, S. (2015, March). Forget psychological science. *Skeptical Inquirer, 39*(2), 9–10.

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and Southern culture of violence effects. *Journal of Personality and Social Psychology, 70*, 740–756.

Angier, N. (2007). *The canon: A whirligig tour of the beautiful basics of science.* New York: Mariner Books.

Appelbaum, B. (2015, September 15). Behaviorists show the US how to nudge. *New York Times*, p. B1, B10.

Ariely, D. (2008). *Predictably irrational.* New York: HarperCollins.

Ariely, D. (2013). *The honest truth about dishonesty: How we lie to everyone–especially ourselves*. Harper Perennial.

Ariely, D. (2015, August 1). The ordeal that made me an observer. *Wall Street Journal*, p. C3.

Ariely, D. (2016, February 6). Office experiments. *Wall Street Journal*, p. C4.

Author. (2016, November). New government reports showcase behavioral science. *APS Observer, 29*, 12.

Azar, B. (1999, November). Crowder mixes theories with humility. *APA Monitor*, p. 18.

Baker, T. B., McFall, R. M., & Shoham, V. (2009). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest, 9*, 67–103.

Banerjee, A., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics, 1*, 151–178.

Barnett, A. (2011, December 20). Is 27 really a dangerous age for famous musicians? A retrospective cohort study. *British Medical Journal.* Retrieved December 28, 2011, from http://www.bmj.com/press-releases/2011/12/20/27-really-dangerous-age-famous-musicians-retrospective-cohort-study

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* New York: Oxford University Press.

Baron, J. (2008). *Thinking and deciding* (4th ed.). Cambridge, MA: Cambridge University Press.

Baron-Cohen, S. (2008). *Autism and Asperger Syndrome: The facts.* Oxford: Oxford University Press.

Beck, D. M. (2010). The appeal of the brain in the popular press. *Perspectives on Psychological Science, 5*, 762–766.

Beck, M. (2008, November 4). And you thought the

Arkes, H. R., & Gaissmaier, W. (2012). Psychological research and the prostate-cancer screening controversy. *Psychological Science, 23*, 547–553.
Arum, R., & Roksa, J. (2011). *Academically adrift*. Chicago: University of Chicago Press.
Asimov, I. (1989). The relativity of wrong. *Skeptical Inquirer, 14,* 35–44.
Associated Press (2007, April 24). Can aspirin prevent cancer? The answer is not clear. *The Washington Post.*
Associated Press (2010, September 7). Safety board renews call for young to fly in own seats. *Wall Street Journal*, A14.
Attari, S. Z., DeKay, M. L., Davidson, C. I., & Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences, 107,* 16054–16059.
Author. (2013, October 19). Trouble at the lab. *The Economist*, pp. 26–30.
debate over fluoridation was settled. *Wall Street Journal*, p. D1.
Beck, M. (2014, September 15). It's time to rethink early cancer detection. *Wall Street Journal*, p. R1–R2.
Begley, S. (2009, October 12). Ignoring the evidence: Why do psychologists reject science? *Newsweek*, p. 30.
Begley, S. (2011, July). The best medicine. *Scientific American*, pp. 50–55.
Bem, D. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100,* 1–19.
Benedetti, F., Carlino, E., & Pollo, A. (2011). How placebos change the patient's brain. *Neuropsychopharmacology, 36,* 339–354.
Berezow, A., & Campbell, H. (2012). *Science left behind: Feel-good fallacies and the rise of the anti-scientific left*. New York: Public Access.

Bertrand, M., Goldin, C., & Katz, L. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics, 2*, 228–255.
Birnbaum, M. H. (1999). Testing critical properties of decision making on the internet. *Psychological Science, 10,* 399–407.
Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55,* 803–832.
Bjorklund, D., & Causey, K. (2017). *Children's thinking: Cognitive development and individual differences* (6th ed.). Thousand Oaks, CA: Sage.
Black, D., Haviland, A., Sanders, S., & Taylor, L. (2008). Gender wage disparities among the highly educated. *Journal of Human Resources, 43*, 630–659.
Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram.* New York: Basic Books.
Blastland, M., & Dilnot, A. (2009). *The numbers game: The commonsense guide to understanding numbers in the news, in politics, and in life*. New York: Gotham Books.
Bloom, P., & Weisberg, D. S. (2007). Childhood origins of adult resistance to science. *Science, 306,* 996–997.
Bluming, A., & Tavris, C. (2009). Hormone replacement therapy: Real concerns and false alarms. *The Cancer Journal, 15,* 93–104.
Boehm, J. K., & Kubzansky, L. (2012). The heart's content: The association between positive psychological well-being and cardiovascular health. *Psychological Bulletin, 138*, 655–691.
Bogle, J. (2015, August). Has the index fund won? *Money Magazine*, pp. 74–77.
Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013).
Brody, J. (2015, August 11). Not vaccinating is the greater risk. *New York Times*, p. D5.
Bronfenbrenner, U., & Mahoney, M. (1975). The structure and verification of hypotheses. In U. Bronfenbrenner & M. Mahoney (Eds.), *Influence on human development.* Hinsdale, IL: Dryden.
Bronowski, J. (1956). *Science and human values.* New York: Harper & Row.
Bronowski, J. (1973). *The ascent of man.* Boston: Little, Brown.
Bronowski, J. (1974). Science, poetry, and human specificity. *American Scholar, 43,* 386–404.
Bronowski, J. (1977). *A sense of the future.* Cambridge, MA: MIT Press.
Bronowski, J. (1978a). *The common sense of science.* Cambridge, MA: Harvard University Press.
Bronowski, J. (1978b). *Magic, science, and civilization.* New York: Columbia University Press.
Bronson, P., & Merryman, A. (2009). *Nurtureshock.* New York: Twelve.
Brooks, A. C. (2008). *Gross national happiness.* New York: Basic Books.
Brown, M. (2010). *How I killed Pluto and why it had it coming*. New York: Spiegel & Grau.
Brownell, K. D. (2011, July). Is there the courage to change America's diet? *APS Observer, 24,* 15–16.
Bryan, C., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, 12653–12656.
Buchtel, E. E., & Norenzayan, A. (2009). Thinking across cultures: Implications for dual processes. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond.* Oxford: Oxford

The pervasive problem with placebos in psychology. *Perspectives on Psychological Science, 8*, 445–454.
Boudry, M., & Buekens, F. (2011). The epistemic predicament of a pseudoscience: Social constructivism confronts Freudian psychoanalysis. *Theoria, 77*, 159–179.
Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* Oxford: Oxford University Press.
Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 23*, 27–34.
Braver, S., Thoemmes, F., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9*, 333–342.
Brewer, M. B. (2013). 25 years toward a multilevel science. *Perspectives on Psychological Science, 8*, 554–555.
Brody, J. E. (2012, January 24). Dental exam went well? Thank fluoride. *New York Times*, p. D7.
University Press.
Buckley, C. (2010, December 9). To test housing program, some are denied aid. *New York Times*, pp. A1–A4.
Bureau of Labor Statistics (2014, June). *How the government measures unemployment*. Washington, DC: Current Population Survey Technical Documentation.
Bureau of Labor Statistics (2017, March 10). Economic news release: Alternative measures of labor underutilization (Table A-15). Retrieved March 12, 2017, from https://www.bls.gov/news.release/empsit.t15.htm
Burrows, G. (2015). *This book won't cure your cancer*. New York: NGO Media.
Burton, R. (2008). *On being certain.* New York: St. Martin's Press.
Bushman, B. J., Baumeister, R. F., Thomaes, S., Ryu, E., Begeer, S., & West, S. (2009). Looking again, and harder, for a link between low self-esteem and aggression. *Journal of Personality, 77*, 427–446.
Bushman, B. J., Anderson, C. A., Donnerstein, E. I., Hummer, T. A., & Warburton, W. (2016). Reply to

comments on SPSSI research summary on media violence by Cupit (2016), Gentile (2016), Glackin & Gray (2016), Gollwitzer (2016), and Krahé (2016). *Analyses of Social Issues and Public Policy, 16*(1), 443–450.
Bushman, B. J., Newman, K., Calvert, S. L., Downey, G., Dredze, M., Gottfredson, M., Jablonski, N. G., Masten, A., Morrill, C., Neill, D. B., Romer, D., & Webster, D. (2016). Youth violence: What we know and what we need to know. *American Psychologist, 71*, 17–39.
Buss, D. M. (2011). *Evolutionary psychology: The new science of the mind* (4th ed.). Boston: Allyn and Bacon.
Cacioppo, J. T. (2007). The structure of psychology. *APS Observer, 20*, 3, 50.
Caldwell, C. (2016, Fall). The hidden costs of immigration. *Claremont Review of Books, 16*(4), 47–50.
Calvert, S., Appelbaum, M. I., Dodge, K. A., Graham, S., Hall, G., Hamby, S. et al. (2017). The American Psychological Association task force assessment of violent video games. *American Psychologist, 72*, 126–143.
Card, N. A. (2011). *Applied meta-analysis for social science research*. New York: Guilford.
Carnagey, N. L., Anderson, C. A., & Bartholow, B. D. (2007). Media violence and social neuroscience: New questions and new opportunities. *Current Directions in Psychological Science, 16,* 178–182.
Carnoy, M., Jacobsen, R., Mishel, L., & Rothstein, R. (2005). *The charter school dust-up: Examining the evidence on enrollment and achievement.* New York: Teachers College Press.
Cartwright, J. (2016). *Evolution and human behavior* (3rd ed.). London: Macmillan.
Cheng, M. (2013, May 20). Measles surges in UK years after vaccine scare. *Yahoo News*. Retrieved January 27, 2017 from https://www.yahoo.com/news/measles-surges-uk-years-vaccine-scare-100011003.html?ref=gs
Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics, 129*, 1553–1623.
Chida, Y., & Hamer, M. (2008). Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: A quantitative review of 30 years of investigations. *Psychological Bulletin, 134,* 829–885.
Cho, H. J., Hotopf, M., & Wessely, S. (2005). The placebo response in the treatment of chronic fatigue syndrome: A systematic review and meta-analysis. *Psychosomatic Medicine, 67,* 301–313.
Churchland, P. M. (1988). *Matter and consciousness* (Rev. ed.). Cambridge, MA: MIT Press.
Churchland, P. S. (2015, August 8). Brain chemicals explain the power of placebos. *Wall Street Journal*, p. C2.
Claridge, G., Clark, K., Powney, E., & Hassan, E. (2008). Schizotypy and the Barnum Effect. *Personality and Individual Differences, 44,* 436–444.
Claypool, H., Hall, C., Mackie, D., & Garcia-Marques, T. (2008). Positive mood, attribution, and the illusion of familiarity. *Journal of Experimental Social Psychology, 44*, 721–728.
Cohen, A. (2008, December 29). Four decades after Milgram, we're still willing to inflict pain. *New York Times*, p. A22.
Collier, L. (2014, July). Defending animal research.

Carvalho, C., Caetano, J.M., Cunha, L., Rebouta, P., Kaptchuk, T.J., & Kirsch, I. (2016). Open-label placebo treatment in chronic low back pain: a randomized controlled trial. *Pain, 157*, 2766–2772.

Castleman, B. (2015). *The 160-character solution*. Baltimore: Johns Hopkins University Press.

Chabris, C., & Simons, D. (2013, March 10). Does this ad make me look fat? *New York Times*, p. SR12.

Chaker, A. (2016, October 26). Are our cars driving us to distraction? *Wall Street Journal*, p. D1, D3.

Chamberlin, J. (2010, November). Type cast. *APA Monitor*, pp. 28–30.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science, 24*, 140–149.

Chen, A. (2014, December 1). At this event, there's madness in the scientific method. *Wall Street Journal*, p. 1, 16.

Cheng, E. (2017, February 11). The logic of my fear of flying. *Wall Street Journal*, p. C4

APA Monitor on Psychology, 44, 40–43.

Conard, E. (2016). *The upside of inequality*. New York: Penguin.

CONSAD Research Corporation (2009, January 12). An analysis of the reasons for the disparity in wages between men and women. *U.S. Department of Labor*, Contract Number GS-23F-02598.

Cook, J. (2016, July). A skeptical response to science denial. *Skeptical Inquirer, 40*(4), 55–57.

Cozby, P. C. (2014). *Methods in behavioral research* (12th ed.). New York: McGraw-Hill.

Craven McGinty, J. (2016, March 5). What the unemployment rate really shows. *Wall Street Journal*, p. A2.

Crawford, J. T. (2012). The ideologically objectionable premise model: Predicting biased political judgments on the left and right. *Journal of Experimental Social Psychology, 48*, 138–151.

Croskerry, P. (2013). From mindless to mindful practice—cognitive bias and clinical decision making. *New England Journal of Medicine, 368*, 2445–2448.

Croswell, J. et al. (2009). Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Annals of Family Medicine, 7*, 212–222.

Cunningham, A., & Zibulsky, J. (2014). *Book smart*. New York: Oxford University Press.

Currier, J. M., Neimeyer, R., & Berman, J. (2008). The effectiveness of psychotherapeutic interventions for bereaved persons: A comprehensive quantitative review. *Psychological Bulletin, 134*, 648–661.

Cuthbertson, A. (2016, March 31). Electric shocks help dyslexic children read faster. *Newsweek*. Retrieved March 2, 2017, from http://www.newsweek.com/electric-shocks-help-dyslexic-children-read-faster-442693

Cuzick, J. (2015, March 23). The evidence is clear it reduces deaths from cancer. *Wall Street Journal*, p. R3.

Dacey, A. (2008). *The secular conscience.* Amherst, NY: Prometheus Books.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*, 512–520.

Davies, J. (2012). Academic obfuscations: The psychological attraction of postmodern nonsense. *Skeptic Magazine, 17*(4). http://www.skeptic.com/eskeptic/14-01-29/#feature

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

Dawkins, R. (2010). *The greatest show on earth*. New York: Free Press.

Dawkins, R. (2016). *The selfish gene: 40th anniversary edition*. New York: Oxford University Press.

Deangelis, T. (2010, November). Getting research into the real world. *APA Monitor*, pp. 60–65.

Deary, I. J. (2013). Intelligence. *Current Biology, 23*,

Demetriou, A., Kui, Z. X., Spanoudis, G., Christou, C., Kyriakides, L., & Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or universal? *Intelligence, 33*, 109–141.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.

DeSoto, K. (2016, March). Under the hood of Mechanical Turk. *APS Observer, 29,* 17–19.

Diacu, F. (2012, October 27). Is failure to predict a crime? *New York Times*, p. A19.

Dingfelder, S. F. (2006, December). Nix the tics. *APA Monitor,* pp. 18–20.

Dingfelder, S. F. (2007, April). Introduction to science. *APA Monitor, 38,* pp. 24–26.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35,* 125–129.

Dodge, K. A., & Rutter, M. (2011). *Gene-environment interactions in developmental psychopathology*. New York: Guilford Press.

Dokoupil, T. (2007, July 16). Trouble in a black box: Did effort to reduce teen suicides backfire? *Newsweek*, p. 48.

Dorlo, P., Betz, W., & Renckens, C. (2015). WHO's strategy on traditional and complementary medicine. *Skeptical Inquirer, 39*(3), 42–45.

Driessen, E., Hollon, S. D., Bockting, C. L. H., Cuijpers, P., & Turner, E. H. (2015). Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US National Institutes of Health-funded trials. *PLoS ONE 10*(9), e0137864. doi:10.1371/journal.pone.0137864

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim,

R673-676.
Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Neuroscience, 11,* 201–211.
DeBakcsy, D. (2014, May). Stop Heisenberg abuse! *Skeptical Inquirer, 38*(3), 40-43.
deCharms, R. C., Maeda, F., Glover, G., Ludlow, D., Pauly, J., Soneji, D., Gabrieli, J., & Mackey, S. (2005). Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences, 102,* 18626–18631.
Deer, B. (2011, January 5). How the case against the MMR vaccine was fixed. *British Medical Journal, 342,* c5347.
de la Cruz, D. (2007, January 5). FTC fines diet-drug marketers. *The Oregonian* (from Associated Press), p. A10.
DeLoache, J., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., Strouse, G., & ODoherty, K. (2010). Do babies learn from baby media? *Psychological Science, 21,* 1570–1574.
L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences, 38*, 1–58, e130. doi:10.1017/S0140525X14000430
Duflo, A., & Karlan, D. (2016, January 1). What data can do to fight poverty. *New York Times*, p. SR12.
Dufresne, T. (Ed.). (2007). *Against Freud: Critics talk back.* Stanford, CA: Stanford University Press.
Duncan, J. (2010). *How intelligence happens*. New Haven, CT: Yale University Press.
Durso, F. T., Nickerson, R. S., Dumais, S., Lewandowsky, S., & Perfect, T. (2007). *Handbook of applied cognition.* Hoboken, NJ: Wiley.
Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading Panel's meta-analysis. *Review of Educational Research, 71,* 393–447.
Eisenberg, L. (1977). The social imperatives of medical research. *Science, 198,* 1105–1110.

Ellenberg, J. (2014). *How not to be wrong*. New York: Penguin Press.
Ellenberg, J. (2015, June 27). The tricks of lying with data. *Wall Street Journal*, p. c3.
Ellis, C. D. (2016). *Index revolution*. Hoboken, NJ: Wiley.
Emery, R. E., Otto, R. K., & O'Donohue, W. T. (2005). A critical assessment of child custody evaluations. *Psychological Science in the Public Interest, 6*(1), 1–29.
Engel, J. (2008). *American therapy*. New York: Gotham Books.
Epley, N. (2013, March 28). What's the question about your field that you dread being asked? *Edge.org: The Reality Club*, Retrieved March 8, 2017, from https://www.edge.org/conversation/whats-the-question-about-your-field-that-you-dread-being-asked
Evans, G., Li, D., & Whipple, S. (2013). Cumulative risk and child development. *Psychological Bulletin, 139,* 1342–1396.
Evans, J. St. B. T. (2015). *How to be a researcher: A strategic guide for academic success* (2nd Ed.). London: Routledge.
Fackelman, K. (1996, November 9). Gastrointestinal blues. *Science News, 150,* 302–303.
Fairless, T. (2017, January 9). Why wages are lagging behind jobs growth. *Wall Street Journal*, p. A2.
Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PloS one, 5*(4). http://dx.doi.org/10.1371/journal.pone.0010068
Ferguson, C. J. (2009). Is psychological research really as good as medical research? *Review of General Psychology, 13,* 130–136.
Ferguson, C. J. (2013). Violent video games and the supreme court. *American Psychologist, 68,* 57–74.
Fischer, P. et al. (2011a). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin, 137,* 517–537.
Fischer, P., Greitemeyer, T., Kastenmüller, A., Vogrincic, C., & Sauer, A. (2011b). The effects of risk-glorifying media exposure on risk-positive cognitions, emotions, and behaviors: A meta-analytic review. *Psychological Bulletin, 137,* 367–390.
Fischhoff, B., & Kadvany, J. (2011). *Risk: A very short introduction*. New York: Oxford University Press.
Flint, J., & Alpert, L. (2016, November 10). Media face backlash for getting it wrong. *Wall Street Journal*, pp. B1–B2.
Foa, E. B., Gillihan, S. J., & Bryant, R. A. (2013). Challenges and successes in dissemination of evidence-based treatments for posttraumatic stress: Lessons learned from prolonged exposure therapy for PTSD. *Psychological Science in the Public Interest, 14*, 65–111.
Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Deknijff, P., Renemieremet, J., Zerjal, T., & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature, 396,* 27–28.
Foster, R. G., & Roenneberg, T. (2008). Human responses to the geophysical daily, annual and lunar cycles. *Current Biology, 18,* R784–R794.
Frank, R. H. (2007). *The economic naturalist*. New York: Basic Books.
Franklin, J., Ribeiro, J., Fox, K., Bentley, K. Kleiman, E., Huang, X., Musacchio, K., Jaroszewski, A., Chang, B., & Nock, M. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin, 143,* 187–232.

Ferguson, C. J. (2015). "Everybody knows psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist, 70*, 527–542.
Ferguson, C. J., Brown, J., & Torres, A. (2017). Education or indoctrination? The accuracy of introductory psychology textbooks in covering controversial topics and urban legends about psychology. *Current Psychology, 36*(1), 1–9.
Feshbach, S., & Tangney, J. (2008). Television viewing and aggression. *Perspectives on Psychological Science, 3*, 387–389.
Firestein, S. (2016). *Failure: Why science is so successful*. New York: Oxford University Press.
Fischer, I., & Savranevski, L. (2015). Extending the two faces of subjective randomness: From the gambler's and hot-hand fallacies toward a hierarchy of binary sequence perception. *Memory & Cognition, 43*, 1056–1070.
Frazier, K. (2015). Facilitated communication has returned from the dead. *Skeptical Inquirer, 39*(3), 9.
Freedman D. (2012, June). The perfected self. *The Atlantic*, pp. 43–52.
Froelich, A., Duckworth, W., & Culhane, J. (2009). Does your iPod really play favorites? *American Statistician, 63*, 263–268.
Furedi, F. (2017). *What's happened to the university?* New York: Routledge.
Furuya-Kanamori, L., & Doi, S. (2016). Angry birds, angry children, and angry meta-analysts: A reanalysis. *Perspectives on Psychological Science, 11*, 408–414.
Gains, P. (2001, February 26). Unravelling dyslexia's riddle. *National Post* (Canada), pp. D1–D2.
Gaissmaier, W., & Gigerenzer, G. (2012). 9/11, Act II: A fine-grained analysis of regional variations in traffic fatalities in the aftermath of the terrorist attacks. *Psychological Science, 23*, 1449–1454.
Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate

psi. *Journal of Personality and Social Psychology, 103*, 933–948.
Gallistel, C. R. (2016, March). The minimum description length principle. *APS Observer, 29*, 5–6.
Galovski, T. E., Malta, L. S., & Blanchard, E. B. (2006). *Road rage: Assessment and treatment of the angry, aggressive driver.* Washington, DC: American Psychological Association.
Gardner, D. (2008). *The science of fear*. New York: Dutton.
Gardner, M. (2006). The memory wars. *Skeptical Inquirer, 30*(1), 28–31.
Gardner, M. (2010). Oprah Winfrey: Bright (but gullible) billionaire. *Skeptical Inquirer, 34*(2), 54–56.
Gaynor, S. T. (2004). Skepticism and caricatures: B. F. Skinner turns 100. *Skeptical Inquirer, 28*(1), 26–29.
Geary, D. C. (2005). *The origin of the mind: Evolution of brain, cognition, and general intelligence.* Washington, DC: American Psychological Association.
Geary, D. C. (2009). *Male, female: The evolution of human sex differences.* Washington, DC: American Psychological Association.
Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*, 847–857.
Gernsbacher, M. A. (2007, May). The value of undergraduate training in psychological science. *APS Observer, 20*, 5–6.
Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.
Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics.
Gorman, C., Cuadros, P., Land, G., Scully, S., & Song, S. (2003, July 28). The new science of dyslexia. *Time*, pp. 53–59.
Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin, 127*, 45–86.
Gosling, S. D., Simine, V., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93–104.
Goswami, U. (Ed.). (2013). *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed.). London: Wiley-Blackwell.
Gould, S. J. (1987). Justice Scalia's misunderstanding. *Natural History, 96*, 14–21.
Grady, D. (2008, June 24). From a prominent death, some painful truths. *New York Times*, p. D5.
Grady, D. (2009, January 6). Should patients be told of better care elsewhere? *New York Times*, p. D1.
Grant, A., & Hofmann, D. (2011). It's not all about me: Motivating hand hygiene among health care professionals by focusing on patients. *Psychological Science, 22*, 1494–1499.
Grant, J. (2011). *Denying science*. Amherst, NY: Prometheus Books.
Grant, K., Sandman, C., Wing, D., Dmitrieva, J., & Davis, E. (2015). Prenatal programming of postnatal susceptibility to memory impairments: A developmental double jeopardy. *Psychological Science, 26*, 1054–1062.
Greene, J. D. (2013). *Moral tribes*. New York: Penguin Press.
Griffin, S., Regnier, E., Griffin, P., & Huntley, V. (2007). Effectiveness of fluoride in preventing cares in

doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*, 53–96.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science, 351*(6277), 1037-b.

Gilovich, T., & Ross, L. (2015). *The wisest one in the room*. New York: Free Press.

Gladwell, M. (2004, December 13). The picture problem. *The New Yorker*, pp. 74–81.

Gladwell, M. (2010, May 17). The treatment. *The New Yorker*, pp. 69–77.

Gleitman, H. (1981). *Psychology.* New York: W. W. Norton.

Goldacre, B. (2008). *Bad science.* London: Fourth Estate.

Gonon, F., Konsman, J., Cohen, D., & Boraud, T. (2012). Why most biomedical findings echoed by newspapers turn out to be false: The case of attention deficit hyperactivity disorder. *PLoS ONE, 7*, e44275. doi:10.1371/journal.pone.0044275

Gopnik, A. (2014, April 5). Want a mind meld? Tell a compelling story. *Wall Street Journal*, p. c2.

Effectiveness of fluoride in preventing caries in adults. *Journal of Dental Research, 86*, 410–415.

Groopman, J. (2007). *How doctors think.* Boston: Houghton Mifflin.

Groopman, J. (2009, November 5). Diagnosis: What doctors are missing. *New York Review of Books*, pp. 26–28.

Haack, S. (2007). *Defending science—within reason: Between scientism and cynicism.* Buffalo, NY: Prometheus Books.

Hacohen, M. C. (2000). *Karl Popper: The formative years, 1902–1945.* Cambridge, England: Cambridge University Press.

Hagen, L. K. (2012). Speechless. *Skeptic Magazine, 17*(3), 14–19.

Haidt, J. (2006). *The happiness hypothesis.* New York: Basic Books.

Hall, H. (2013). "New study shows…" *Skeptic Magazine, 18*(2), 4–5.

Hall, H. (2016). Clear thinking about cancer. *Skeptical Inquirer, 40*(2), 57–59.

Hand, D. (2014). *The improbability principle*. New York: Farrar, Straus, & Giroux.

Hariri, A. R., & Holmes, A. (2006). Genetics of emotional regulation: The role of the serotonin transporter in neural function. *Trends in Cognitive Sciences, 10*, 182–191.

Harlow, H. F., & Suomi, S. J. (1970). The nature of love—Simplified. *American Psychologist, 25*, 161–168.

Harriet, H. (2008, September/October). "We couldn't say it in print if it wasn't true." *Skeptical Inquirer, 32*, 46–49.

Harrington, A. (2008). *The cure within.* New York: Norton.

Hassani, S. (2016, July). Does E=mc^2 imply mysticism? *Skeptical Inquirer, 40*(4), 46–49.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world.* Thousand Oaks, CA: Sage.

Heinzen, T., Lilienfeld, S., & Nolan, S. (2014). *The horse that won't go away: Clever Hans, facilitated communication, and the need for clear thinking*. New York: Worth Publishers.

Hemingway, M. (2016, November 21). Things poll apart. *Weekly Standard*, pp. 20–22.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 1–75.

Hewer, M. (2014, December). Selling sweet nothings. *APS Observer, 27*, 14–18.

Hines, T. M. (2003). *Pseudoscience and the paranormal* (2nd ed.). Buffalo, NY: Prometheus Books.

Holton, G., & Roller, D. (1958). *Foundations of modern physical science*. Reading, MA: Addison-Wesley.

Honda, H., Shimizu, Y., & Rutter, M. (2005). No effect of MMR withdrawal on the incidence of autism: A total population study. *Journal of Child Psychology and*

Insurance Institute for Highway Safety (2005, July 16). If you drive while phoning you're far more likely to get into a crash in which you'll be injured. *Status Report, 40*(6), 1–3.

Investor's Guide (2017, January). The fund report. *Money Magazine*, p. 108.

Isaacson, W. (2011). *Steve Jobs*. New York: Simon & Schuster.

Jaffe, E. (2005). How random is that? Students are convenient research subjects but they're not a simple sample. *APS Observer, 18*(9), 20–30.

Jaffe, E. (2011, September). Identity shift. *APS Observer, 24*, 12–16.

Jaffe, E. (2012, February). Rewired: Cognition in the digital age. *APS Observer, 25*, 16–20.

Jaffee, S., Strait, L., & Odgers, C. (2012). From correlates to causes. *Psychological Bulletin, 138*, 272–295.

Johnson, S. (2007). *The ghost map.* New York: Riverhead Books.

Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology, 66*, 116–133.

Kachka, B. (2012, November 5). Proust wasn't a neuroscientist. *New York Magazine*, pp. 30–33, 90–92.

Kagan, J. (2006). *An argument for mind.* New Haven, CT: Yale University Press.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8*, 407–424.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.

Kalat, J. W. (2007). *Biological psychology* (9th ed.).

population study. *Journal of Child Psychology and Psychiatry, 46,* 572–579.

Hood, B. (2017, March 12). No evidence to back idea of learning styles. *The Guardian*. Retrieved March 13 from https://www.theguardian.com/education/2017/mar/12/no-evidence-to-back-idea-of-learning-styles?CMP=twt_a-science_b-gdnscience

Horswill, M. (2016). Hazard perception in driving. *Current Directions in Psychological Science, 25*, 425–430.

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7*, 1–5.

Hung, E. (2013). *Philosophy of Science Complete*. Boston: Wadsworth.

Hunt, E. (2011). *Human intelligence*. Cambridge, MA: Cambridge University Press.

Immen, W. (1996, August 8). Could you repeat that in Klingon? *Globe & Mail (Toronto).*

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7*, 496–503.

Kalat, J. W. (2007). *Biological psychology* (9th ed.). Belmont, CA: Wadsworth.

Kalanithi, P. (2016). *When breath becomes air*. New York: Random House.

Kalb, C., & White, E. (2010, May 24). What should you really be afraid of? *Newsweek*, p. 64.

Kantrowitz B., & Underwood A. (1999, November 22). Dyslexia and the new science of reading. *Newsweek*, pp. 72–78.

Kaplan, M., & Kaplan, E. (2007). *Chances are: Adventures in probability*. New York: Penguin Books.

Keating, D. P. (2007). Understanding adolescent development: Implications for driving safety. *Journal of Safety Research, 38,* 147–157.

Kenney, C. (2008). *The best practice: How the new quality movement is transforming medicine.* New York: PublicAffairs Books.

Kholodkov, T. (2013, February). Untangling the web. *APS Observer, 26,* 36–37.

King, B. M. (2013). The modern obesity epidemic, ancestral hunter-gatherers, and the sensory/reward

control of food intake. *American Psychologist, 68,* 88–96.

Kirp, D. (2017, January 8). Text your way to college. *New York Times*, p. 12.

Kirschner, P., & van Merriënboer, J. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*, 169–183.

Klatzky, R. L. (2012, November). When it comes to department name, “psychology” is #1. *APS Observer, 25,* 8–9.

Klein, D. N. (2010). Chronic depression: Diagnosis and classification. *Current Directions in Psychological Science, 19*, 96–100.

Koepsell, D. (2015, November). From N-rays to EmDrives: When does science become pseudoscience? *Skeptical Inquirer, 39*(6), 14–15.

Kolata, G. (2009, August 20). One injured hamstring, a string of treatments. *New York Times*, p. E8.

Kolata, G. (2014, February 12). Vast study casts doubts on value of mammograms. *New York Times*, p. A1, A3.

Kolata, G. (2016a, May 25). Latest trend in treating prostate cancer: Don’t. *New York Times*, p. A1.

Kolata, G. (2016b, December 13). No single answer. *New York Times*, pp. D1–D2.

Kolata, G. (2017, March 23). A scholarly sting operation shines a spotlight on “predatory” scientific journals. *New York Times*, p. A19.

Kolbert, E. (2005, September 19). Storm warnings. *The New Yorker*, pp. 35–36.

Kolesnikova, N., & Liu, Y. (2011, October). Gender wage gap may be much smaller than most think. *The Regional Economist*. Federal Reserve Bank of St. Louis. Retrieved February 14, 2017, from https://www.

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review, 15,* 1135–1140.

Kushner, H. I. (1999). *A cursing brain? The histories of Tourette Syndrome.* Cambridge, MA: Harvard University Press.

Landsburg, S. E. (2007). *More sex is safer sex: The unconventional wisdom of economics*. New York: Free Press.

Larrick, R. P., Timmerman, T. A., Carton, A. M., & Abrevaya, J. (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science, 22*, 423–428.

Layton, L., & Koh, E. (2015, July 14). One of today’s great generation gaps: Whether or not you think Pluto is a planet. *Washington Post*. Retrieved February 14, 2017, from https://www.washingtonpost.com/local/education/one-of-todays-great-generation-gaps-whether-or-not-you-think-pluto-is-a-planet/2015/07/14/7dcb4f6e-2998-11e5-a5ea-cf74396e59ec_story.html?utm_term=.3d1f009a5b0d

Lazarsfeld, P. (1949). The American Soldier—an expository review. *Public Opinion Quarterly, 13,* 377–404.

Lemann, N. (2012, April 23). Evening the odds. *New Yorker*, pp. 69–73.

Leonhardt, D. (2017, March 7). A public-health crisis that we can fix. *New York Times*, p. A27.

Levitin, D. J. (2016). *A field guide to lies*. New York: Dutton.

Levy, E. (2009, February 13). The maggots in your mushrooms. *New York Times*, p. A25.

Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving: Is there any stopping the psychological

stlouisfed.org/Publications/Regional-Economist/October-2011/Gender-Wage-Gap-May-Be-Much-Smaller-Than-Most-Think?hc_location=ufi#endnotes

Korownyk, C. et al. (2014). Televised medical talk shows—what they recommend and the evidence to support their recommendations: A prospective observational study. *BMJ, 349*, g7346. doi:10.1136/bmj.g7346

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences. *American Psychologist, 70*, 543–556.

Kosova, W., & Wingert, P. (2009, June 8). Crazy talk. *Newsweek*, pp. 54–62.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology, 36*, 153–159.

Krueger, J. I., Vohs, K. D., & Baumeister, R. F. (2008). Is the allure of self-esteem a mirage after all? *American Psychologist, 63*, 64–65.

Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology, 88*, 725–735.

refractory period? *Psychological Science, 17*, 228–235.

Levy, S. (2005, January 31). Does your iPod play favorites? *Newsweek*, p. 10.

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science, 24*, 622–633.

Lewis, M. (2004). *Moneyball*. New York: Norton.

Lewis, M. (2017). *The undoing project*. New York: Norton.

Li, C. (1975). *Path analysis: A primer.* Pacific Grove, CA: Boxwood Press.

Lilienfeld, S. O. (2005). The 10 commandments of helping students distinguish science from pseudoscience in psychology. *APS Observer, 18*(9), 39–51.

Lilienfeld, S. O. (2006, February). Correlation still isn't causation. *APS Observer, 19*, 9.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70.

Lilienfeld, S. O. (2010). Can psychology become a science? *Personality and Individual Differences, 49*, 281–288.

Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist, 67*, 111–129.

Lilienfeld, S. O. (2013, March). Closing the science-practice gap. *APS Observer, 26*, 8–11.

Lilienfeld, S. O. (2014, July). Using pseudoscience to shine light on good science. *APS Observer, 27*, 33–34.

Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science, 12*, 138–169.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50*, 7–36.

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2010). *50 great myths of popular psychology*. Malden, MA: Wiley-Blackwell.

Lilienfeld, S. O., Ritschel, L., Lynn, S., Cautin, R., & Latzman, R. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science, 9*, 355–387.

Lilienfeld, S. O., Ruscio, J., & Lynn, S. J. (Eds.). (2008). *Navigating the mindfield: A guide to separating science from pseudoscience in mental health.* Buffalo, NY: Prometheus Books.

Lit, L., Schweitzer, J., & Oberbauer, A. (2011). Handler beliefs affect scent detection dog outcomes. *Animal Cognition, 14*, 387–394.

Liu, C., Floud, S., Pirie, K., Green, J., Peto, R., & Beral, V. (2016). Does happiness itself directly affect mortality? psychotherapy: Problems and pitfalls. *Skeptical Inquirer, 27*(4), 40–46.

Magee, B. (1985). *Philosophy and the real world: An introduction to Karl Popper.* LaSalle, IL: Open Court.

Maizels, M. (2005). Why should physicians care about behavioral research? *Headache, 45*, 411–413.

Majima, Y. (2015). Belief in pseudoscience, cognitive style and science literacy. *Applied Cognitive Psychology, 29*, 552–559.

Malkiel, B. G. (2016). *A random walk down Wall Street*. New York: Norton.

Maniaci, M. R., & Rogge, R. D. (2014). Conducting research on the internet. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 443–470). Cambridge, UK: Cambridge University Press.

Manjoo, F. (2008). *True enough: Learning to live in a post-fact society*. Hoboken, NJ: John Wiley.

Marchant, J. (2016, January 10). A placebo treatment for pain. *New York Times*, p. SR5.

Marks, D. F. (2001). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marks, L. (2015, March 23). The risks are large and increase as a person ages. *Wall Street Journal*, p. R3.

Martin, L. et al. (2011). The "distressed" personality, coping and cardiovascular risk. *Stress and Health, 27*, 64–72.

Martin, R., & Hull, R. (2006). The case study perspective on psychological research. In R. J. Sternberg, H. L. Roediger, & D. F. Halpern (Eds.), *Critical thinking in psychology* (pp. 90–109). New York: Cambridge University Press.

Matthews, K. A. (2013). Matters of the heart: Advancing

The prospective UK million women study. *The Lancet, 387*, 874–881.

Loftus, E. F., & Guyer, M. J. (2002, May/June). Who abused Jane Doe: The hazards of the single case history. *Skeptical Inquirer, 26*(3), 24–32.

Lohr, S., & Singer, N. (2016, November 10). How data failed us in an election. *New York Times*, p. B1, B5.

Lowman, R., & Benjamin, L. T. (2012). Psychology and the national medal of science. *American Psychologist, 67*, 174–183.

Lu, S. (2015, April). Great expectations. *APA Monitor on Psychology, 45*, 50–53.

Lukianoff, G. (2012). *Unlearning liberty: Campus censorship and the end of American debate*. New York: Encounter Books.

Lukianoff, G., & Haidt, J. (2015, September). The coddling of the American mind. *The Atlantic.* http://www.theatlantic.com/magazine/archive/2015/2009/the-coddling-of-the-american-mind/399356/

Lynn, S. J., Loftus, E. F., Lilienfeld, S. O., & Lock, T. (2003). Memory recovery techniques in psychological perspectives on cardiovascular diseases. *Perspectives on Psychological Science, 8*, 676–678.

Matute, H., Yarritu, I., & Vadillo, M. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology, 102*, 392–405.

Maxwell, S., Lau, M., & Howard, G. (2015). Is psychology suffering from a replication crisis? What does "Failure to replicate" Really mean? *American Psychologist, 70*, 487–498.

Mazur, J. (2016). *Fluke: The math and myth of coincidence*. New York: Basic Books.

McAdams, T., Neiderhiser, J. M., Rijsdijk, F., Narusyte, J., Lichtenstein, P., & Eley, T. (2014). Accounting for genetic and environmental confounds in associations between parent and child characteristics. *Psychological Bulletin, 140*, 1138–1173.

McCloskey, M. (1983, April). Intuitive physics. *Scientific American, 248*(4), 122–130.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodword, M., Haworth, C., Palamara, P., &

Cercarelli, R. (2005, August 20). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *British Medical Journal, 331*(7514), 428.

McHugh, P. (2008). *The memory wars: Psychiatry's clash over meaning, memory, and mind.* Washington, DC: The Dana Foundation.

McKinley, J., Dempster, M., & Gormley, G. (2015). "Sorry, I meant the patient's left side": Impact of distraction on left–right discrimination. *Medical Education, 49*, 427–435.

McLanahan, S., Tach, L., & Schneider, D. (2013). The causal effects of father absence. *Annual Review of Sociology, 39*, 399–427.

McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest, 4*(2), 45–79.

McNally, R. J., & Geraerts, E. (2009). A new solution to the recovered memory debate. *Perspectives on Psychological Science, 4*, 126–134.

Medawar, P. B. (1967). *The art of the soluble.* London: Methuen.

Medawar, P. B. (1979). *Advice to a young scientist.* New York: Harper & Row.

Medawar, P. B. (1984). *The limits of science.* New York: Harper & Row.

Medawar, P. B. (1990). *The threat and the glory.* New York: HarperCollins.

Medawar, P. B., & Medawar, J. S. (1983). *Aristotle to zoos: A philosophical dictionary of biology.* Cambridge, MA: Harvard University Press.

Medin, D. (2012). "Rigor Without Rigor Mortis: The APS Board Discusses Research Integrity," *Observer*, p. 6.

Meehl, P. E. (1954). *Clinical versus statistical prediction:*

Mischel, W. (2015). *The marshmallow test: Why self-control is the engine of success.* New York: Little, Brown.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*, 109–117.

Mlodinow, L. (2008). *The drunkard's walk: How randomness rules our lives.* New York: Pantheon.

Mole, B. (2016, June 21). Googling medical symptoms may no longer convince you that you're dying. *Ars Technica.* Retrieved March 11, 2017, from https://arstechnica.com/science/2016/06/googling-medical-symptoms-may-no-longer-convince-you-that-youre-dying/

Mook, D. G. (2001). *Psychological research: The ideas behind the methods.* New York: Norton.

Mooney, C. (2005). *The republican war on science*. New York: Basic Books.

Moore, S. A., & Zoellner, L. (2007). Overgeneral autobiographical memory and traumatic events: An evaluative review. *Psychological Bulletin, 133*, 419–437.

Morera, O. F., & Dawes, R. M. (2006). Clinical and statistical prediction after 50 years: A dedication to Paul Meehl. *Journal of Behavioral Decision Making, 19*, 409–412.

Moskowitz, T., & Wertheim, L. (2011). *Scorecasting*. New York: Crown.

Mostofsky, E., Penner, E. A., & Mittleman, M. A. (2014). Outbursts of anger as a trigger of acute cardiovascular events: A systematic review and meta-analysis. *European Heart Journal, 35*, 1404–1410.

Murray, C. (2012). *Coming apart*. New York: Crown Forum.

Myers, D. (2015, January). Happy marriages and healthy bodies. *APS Observer, 28*, 36–37.

Myers, D. (2017, February). People need people: Why

A theoretical analysis and review of the literature. Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meyers, L. (2008, January). Recommended reading: Psychologists share the contents of their self-help shelves. *APA Monitor*, pp. 26–27.

Michel, A. (2015, March). Countering neuromyths in the movies. *APS Observer, 28,* 31–32.

Mielczarek, E., & Engler, B. (2012). Measuring mythology: Startling concepts in NCCAM grants. *Skeptical Inquirer, 36*(1), 35–43.

Mielczarek, E., & Engler, B. (2013). Nurturing non-science. *Skeptical Inquirer, 37*(3), 32–39.

Mielczarek, E., & Engler, B. (2014). Selling pseudoscience. *Skeptical Inquirer, 38*(3), 44-51.

Miller, K. R. (2008). *Only a theory: Evolution and the battle for America's soul.* New York: Viking.

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders. *American Psychologist, 61,* 10–26.

close relationships predict health. *APS Observer, 30,* 42–43.

National Academies of Sciences (2016). *Genetically engineered crops: Experiences and prospects*. Washington, DC: The National Academies Press. doi:10.17226/23395

National Highway Traffic Safety Administration (2014). Fatality analysis reporting system. Retrieved January 24, 2017, from https://www-fars.nhtsa.dot.gov/Main/index.aspx

National Reading Panel: Reports of the Subgroups (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC.

National Safety Council (2016). *Injury facts 2016: Odds of dying*. Retrieved January 24, 2017, from http://www.nsc.org/learn/safety-knowledge/Pages/injury-facts-odds-of-dying.aspx

Nattrass, N. (2012). The social and symbolic power of AIDS denialism. *Skeptical Inquirer, 36*(4), 34–38.

Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., & Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine, 11*(1), 222. doi:10.1186/1741-7015-11-222

Nickell, J., & McGaha, J. (2015, November). The search for negative evidence. *Skeptical Inquirer, 39*(6), 53–55.

Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The Annals of the American Academy of Political and Social Science, 658*, 36–66.

Nisbet, M. (2016). Winning the vaccine war. *Skeptical Inquirer, 40*(6), 27–29.

Nolen-Hoeksema, S., Wisco, B., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science, 3,* 400–424.

Novella, S. (2012, May). Pseudoscience in our universities. *Skeptical Inquirer, 36*(3), 24–25.

Novella, S. (2015, May). It's time for science-based medicine. *Skeptical Inquirer, 39*(3), 22–24.

Novotney, A. (2008, July). Custody collaborations. *APA Monitor*, pp. 49–51.

Novotney, A. (2009, February). Dangerous distraction. *APA Monitor*, pp. 32–36.

Nyhan, B., Reifler, J., Richey, S., & Freed, G. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics, 133*(4), 1–8.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin, 133,* 310–327.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*, 15583–15587.

Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171–192.

Otto, S. (2016). *The war on science*. Minneapolis, MN: Milkweed Editions.

Overskeid, G. (2007). Looking for Skinner and finding Freud. *American Psychologist, 62,* 590–595.

Paloutzian, R. F., & Park, C. L. (Eds.). (2005). *Handbook of the psychology of religion and spirituality.* New York: Guilford Press.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188.

Parker, I. (2010, May 17). The poverty lab. *The New Yorker*, pp. 79–89.

Parker, K. (2014, February 28). Obama's best hope for change. *Washington Post*. https://www.washingtonpost.com/opinions/kathleen-parker-obamas-best-hope-for-change/2014/02/28/0a7bc986-a0b8-11e3-a050-dc3322a94fa7_story.html?utm_term=.ff859e3156c2

Parker-Pope, T. (2009, January 13). A problem of the brain, not the hands. *New York Times*, p. D5.

Parker-Pope, T. (2011, October 10). Prostate test finding leaves a swirl of confusion. *New York Times Wellness Blog*, pp. 1–4. Retrieved October 12, 2011, from http://well.blogs.nytimes.com/2011/10/10/

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition, 37,* 632–643.

Observations. (2017, January). Loftus receives 2016 John Maddox Prize. *APS Observer, 30,* 9.

O'Connor, A. (2011, December 27). Really? *New York Times*, p. D5.

Offit, P. A. (2011). *Deadly choices: How the anti-vaccine movement threatens us all*. New York: Basic Books.

Offit, P. A. (2008). *Autism's false prophets.* New York: Columbia University Press.

Olson, R. K. (2004). SSSR, environment, and genes. *Scientific Studies of Reading, 8,* 111–124.

O'Neill, J., & O'Neill, D. (2012). *The declining importance of race and gender in the labor market*. Washington, DC: AEI Press.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943.

prostate-test-finding-leaves-a-swirl-of-confusion/

Parry, M. (2012, July 22). Please be advised. *New York Times Education Life*, p. 24–27.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9,* 105–119.

Patihis, L., Ho, L., Tingen, I., Lilienfeld, S., & Loftus, E. (2014). Are the "memory wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science, 25*, 519–530.

Paulos, J. (2016, March). A numerate life. *Skeptical Inquirer, 40*(2), 49–52.

Peer, E., & Solomon, L. (2012). Professionally biased: Misestimations of driving speed, journey time and time-savings among taxi and car drivers. *Judgment and Decision Making, 7*, 165–172.

Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 453–472). Malden, MA: Blackwell.

Peterson, A. (2013, October 29). Pediatricians set limits for TV, internet, cellphone use. *Wall Street Journal*, p. D1.

Peterson, B. (2012, April 9). It's not a tumor! The psychology behind cyberchondria. *Newsweek*. Retrieved March 11, 2017, from: http://www.newsweek.com/its-not-tumor-psychology-behind-cyberchondria-64071

Peterson, R., & Pennington, B. F. (2012). Developmental dyslexia. *Lancet, 379*, 1997–2007.

Petry, N. M. (2005). *Pathological gambling: Etiology, comorbidity, and treatment.* Washington, DC: American Psychological Association.

Phelps, E. (2013, December). Educating consumers of psychological science. *APS Observer, 26,* 5.

Pigliucci, M. (2010). *Nonsense on stilts*. Chicago: University of Chicago Press.

Pinker, S. (1997). *How the mind works.* New York: W. W. Norton.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.

Pinker, S. (2016, January 2). New advances in behavioral genetics. *Wall Street Journal*, p. C2.

Plante, C., & Anderson, C. (2017, February). Global warming and violent behavior. *APS Observer, 30,* 29–32.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*, 3–23.

Pohl, R. (Ed.). (2017). *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory.* New York: Routledge.

Polidoro, M. (2015, March). Enter the boy wonder: Randi

Rabin, R. (2009, June 16). Alcohol's good for you? Some scientists doubt it. *New York Times*, p. D1.

Radford, B. (2009). Psychic exploits horrific abduction case. *Skeptical Inquirer, 33*(6), 6–7.

Radford, B. (2010). The psychic and the serial killer. *Skeptical Inquirer, 34*(2), 32–37.

Radford, B. (2011). Left brained or right brained. *Skeptical Inquirer, 35*(1), 22.

Radford, B. (2016a). Facilitated communication consent claimed in sexual assault trial. *Skeptical Inquirer, 40*(1), 8.

Radford, B. (2016b). Italian court acquits six convicted seismologists. *Skeptical Inquirer, 40*(3), 9.

Raine, A. (2008). From genes to brain to antisocial behavior. *Current Directions in Psychological Science, 17*, 323–328.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review, 27,* 224–260.

Randall, L. (2005, September 18). Dangling particles. *New York Times*, p. WK13.

Randi, J. (2005). Fakers and innocents: The one million dollar challenge and those who try for it. *Skeptical Inquirer, 29*(4), 45–50.

Randi, J. (2011). Twas brillig: Trying to give away a million dollars. *Skeptic Magazine, 16*(4), 8–9.

Randi, J. (2017, March). The dangerous delusion about vaccines and autism. *Skeptical Inquirer, 41*(2), 29–31.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading*. New York: Psychology Press.

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest, 17*, 4–34.

Reddy, S. (2016, January 12). New advice fails to stem

vs Geller. *Skeptical Inquirer, 39*(2), 33–35.
Popper, K. R. (1959). *The logic of scientific discovery.* New York: Harper & Row.
Popper, K. R. (1963). *Conjectures and refutations.* New York: Harper & Row.
Popper, K. R. (1972). *Objective knowledge.* Oxford, England: Oxford University Press.
Popper, K. R. (1976). *Unended quest: An intellectual biography.* La Salle, IL: Open Court.
Poppy, C. (2017, January). Survey shows Americans fear ghosts, the government, and each other. *Skeptical Inquirer, 41*(1), 16–18.
Powell, J. (2015). *Four revolutions in the earth sciences*. New York: Columbia University Press.
Price, E. (2009, July). Behavioral research can help curb swine flu. *APA Monitor*, p. 11.
Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences, 11*, 37–43.
Quart, A. (2012, November 25). Neuroscience: Under attack. *New York Times*, p. SR14.
confusion on mammograms. *Wall Street Journal*, p. D1–D2.
Reif, L. (2016, December 6). The dividends of funding basic science. *Wall Street Journal*, p. A17.
Reporting Science (2012, September 22). Journalistic deficit disorder. *The Economist*, pp. 90–92.
Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1432–1440.
Richtel, M. (2014). *A deadly wandering*. New York: HarperCollins.
Richtel, M. (2016, May 22). It's no accident: Advocates want to speak of car "crashes" instead *New York Times*. Retrieved February 18, 2017, from https://www.nytimes.com/2016/05/23/science/its-no-accident-advocates-want-to-speak-of-car-crashes-instead.html?_r=0
Riener, C., Proffitt, D. R., & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12*, 740–745.

Rind, B. (2008). The Bailey affair: Political correctness and attacks on sex research. *Archives of Sexual Behavior, 37*, 481–484.
Ritchie, S. (2015). *Intelligence: All that matters*. London: John Murray Learning.
Robles, T., Slatcher, R., Trombello, J., & McGinn, M. (2014). Marital quality and health: A meta-analytic review. *Psychological Bulletin, 140*, 140–187.
Rodriguez, F., Rhodes, R. E., Miller, K. F., & Shah, P. (2016). Examining the influence of anecdotal stories and the interplay of individual differences on reasoning. *Thinking & Reasoning, 22*, 274–296.
Roediger, H. L. (2016). Varieties of fame in psychology. *Perspectives on Psychological Science, 11*, 882–887.
Roller, L., & Gowan, J. (2011). Disease state management: The placebo-nocebo conundrum. *Australian Journal of Pharmacy, 92*, 76–79.
Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 46*, 775–776.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.
Russo, F. (1999, May). The clinical-trials bottleneck. *The Atlantic Monthly*, pp. 30–36.
Salik, B. (2016, March 11). Calculus is still useful in the calculator age. *Wall Street Journal*, p. A14.
Salthouse, T. A. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology, 63*, 201–226.
Sanger-Katz, M. (2014, November 23). Health care myths. *New York Times*, p. 10.
Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience.* New York:
Seethaler, S. (2009). *Lies, damned lies, and science*. Upper Saddle River, NJ: Pearson Education.
Seidenberg, M. (2017). *Language at the speed of sight*. New York: Basic Books.
Seife, C. (2010). *Proofiness: The dark arts of mathematical deception*. New York: Viking.
Seppa, N. (2006, January 14). Put down that fork: Studies document hazards of obesity. *Science News, 169*, 21.
Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology, 73*, 6–14.
Shaffer, R., & Jadwiszczok, A. (2010). Psychic defective: Sylvia Browne's history of failure. *Skeptical Inquirer, 34*(2), 38–42.
Shapin, S. (2006, November 6). Sick city: Maps and mortality in the time of cholera. *The New Yorker*, pp. 110–115.
Shapiro, A., Shapiro, E., Bruun, R., & Sweet, R. (1978). *Gilles de la Tourette syndrome.* New York: Raven Press.
Sharot, T. (2011). *Optimism bias*. New York: Pantheon.
Shaywitz, S. E., & Shaywitz, B. A. (2004). Neurobiologic basis for reading and reading disability. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 417–442). Baltimore: Paul Brookes.
Shermer, M. (2005). *Science friction: Where the known meets the unknown.* New York: Times Books.
Shermer, M. (2011). *The believing brain*. New York: Times Books.
Shermer, M. (2012). The reality distortion field. *Skeptic Magazine, 17*(4), 29–31.
Shermer, M. (2016, August). The quack of the gaps problem. *Scientific American, p. 75*.
Shermer, M. (2017). *Skeptic*. New York: St. Martins

The seductive appeal of mindless neuroscience. New York: Basic Books.
Schaie, K. W., & Willis, S. (Eds.). (2010). *Handbook of the psychology of aging* (7th ed.). San Diego: Academic Press.
Schmidt, F. L., & Oh, I. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology, 4*, 32–37.
Scholl, S. G., & Greifeneder, R. (2011). Disentangling the effects of alternation rate and maximum run length on judgments of randomness. *Judgment and Decision Making, 6,* 531–541.
Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigne, K. C. (2016). The role of neuroscience within psychology: A call for inclusiveness over exclusiveness. *American Psychologist, 71*, 52–70.
Schwarz, K. A., Pfister, R., & Buchel, C. (2016). Rethinking explicit expectations: Connecting placebos, social cognition, and contextual perception. *Trends in Cognitive Sciences, 20*, 469–480.
Scott, E. C. (2005). *Evolution vs. creationism.* Berkeley, CA: University of California Press.
Shermer, M. (2017). *Skeptic*. New York: St. Martins.
Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*, 771–799.
Sielski, M. (2010, November 17). Rethinking quarterback stats. *Wall Street Journal*, p. D6.
Silver, N. (2012). *The signal and the noise*. New York: Penguin Books.
Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22,* 1359–1366.
Simons, D., Boot, W., Charness, N., Gathercole, S., Chabris, C., Hambrick, D., & Stine-Morrow, E. (2016). Do "Brain-training" programs work? *Psychological Science in the Public Interest, 17*, 103–186.
Simonton, D. K. (2015, April). Unifying psychology as a physical science. *APS Observer, 28,* 18–24.
Singh, K., Spencer, A., & Brennan, D. (2007). Effects of water fluoride exposure at crown completion and maturation on caries of permanent first molars. *Caries Research, 41,* 34–42.

Singh, A., Uijtdewilligen, L., Twisk, J., Mechelen, W., & Chinapaw, M. (2012). Physical activity and performance at school. *Archives of Pediatric Adolescent Medicine, 166*, 49–55.
Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist, 91,* 6–7.
Skenazy, L. (2009). *Free-range kids*. San Francisco: Jossey-Bass.
Skitka, L., & Sargis, E. (2006). The internet as psychological laboratory. *Annual Review of Psychology, 57,* 529–555.
Sleek, S. (2013, November). Attacking science: Trials of inconvenient truth tellers. *APS Observer, 26,* 24–29.
Sleek, S. (2015, September). Mischel garners a Golden Goose Award. *APS Observer, 28,* 9.
Sleek, S. (2017, March). The John Maddox Prize nomination for Elizabeth Loftus. *Skeptical Inquirer, 41*(2), 20–23.
Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making, 2,* 79–95.
Slovic, S., & Slovic, P. (2015). *Numbers and nerves: Information, emotion, and meaning in a world of data*. Corvallis, OR: Oregon State University Press.
Smith, P. (2013). *Cockpit confidential*. Naperville, IL: Sourcebooks.
Smith, T., Polloway, E., Patton, J., & Dowdy, C. (2016). *Teaching students with special needs in inclusive settings* (6th ed.). Upper Saddle River, NJ: Pearson Education.
Snyderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political
Sternberg, R. J. (2016). "Am I famous yet?" Judging scholarly merit in psychological science: An introduction. *Perspectives on Psychological Science, 11*, 877–881.
Sternberg, R. J. & Kaufman, S. B. (Eds.). (2011). *Cambridge handbook of intelligence*. New York: Cambridge University Press.
Sternberg, R. J., Roediger, H. L., & Halpern, D. F. (Eds.). (2006). *Critical thinking in psychology.* New York: Cambridge University Press.
Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G. et al. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10*, 479–491.
Stix, G. (2015, Winter). How to build a better learner. *Scientific American Mind Special Collector's Edition, 23*(4), 68–75.
Strayer, D. L., Cooper, J. C., Turrill, J., Coleman, J., & Hopman, R. (2016). Talking to your car can drive you to distraction. *Cognitive Research: Principles and Implications, 1*, 1–16.
Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16,* 128–131.
Strayer, D. L., Medeiros-Ward, N., & Watson, J. (2013). Gender invariance in multitasking. *Psychological Science, 24*, 809–810.
Suski, V., & Stacy, M. (2016). Tic and Tourette Syndrome. In R. Lisak, D. Truong, W. Carroll & R. Bhidayasiri (Eds.), *International Neurology* (pp. 214–216). Chichester, West Sussex, UK: Wiley Blackwell.
Swan, L., Skarsten, S., Heesacker, M., & Chambers, J. (2015). Why psychologists should reject complemen

racism: Problems of motive attribution in political analysis. *Journal of Social Issues*, 129–150.

Solberg, E. & Laughlin, T. (1995). The gender pay gap, fringe benefits, and occupational crowding. *Industrial and Labor Relations Review, 48*, 692–708.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin.* Chicago: University of Chicago Press.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought.* New Haven, CT: Yale University Press.

Stanovich, K. E. (2010). *Decision making and rationality in the modern world*. New York: Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science, 22*, 259–264.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge, MA: MIT Press.

(2015). Why psychologists should reject complementary and alternative medicine. *Professional Psychology: Research and Practice, 46*, 325–339.

Tait, R., Chibnall, J., & Kalauokalani, D. (2009). Provider judgments of patients in pain: Seeking symptom certainty. *Pain Medicine, 10*, 11–34.

Tanaka, H. et al. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science, 22*, 1442–1451.

Taubes, G. (2017). *The case against sugar*. New York: Knopf.

Tavris, C. (2014, October). Teaching contentious classics. *APS Observer, 27*, 12–16.

Taylor, A. K., & Kowalski, P. (2004). Naive psychological science: The prevalence, strength, and sources of misconceptions. *Psychological Record, 54*, 15–25.

Taylor, B. (2006). Vaccines and the changing epidemiology of autism. *Child Care, Health, and Development, 32*, 511–519.

Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology*, 509–529.

Tetlock, P. E. (2012). Rational versus irrational prejudices: How problematic is the ideological lopsidedness of social psychology? *Perspectives on Psychological Science, 7*, 519–521.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting*. New York: Crown.

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. New York: Norton.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

Thornton, E. (1986). *The Freudian fallacy.* London: Paladin Books.

Thomas, K., De Freitas, J., DeScioli, P., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General, 145*, 621–629.

Tilburt, J. C., Emanuel, E. J., Kaptchuk, T. J., Curlin, F. A., & Miller, F. G. (2008, October 23). Prescribing placebo treatments: Results of national survey of US internists and rheumatologists. *BMJ, 337*, a1938. doi:10.1136/bmj.a1938

Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making, 20*, 103–124.

Tracey, T., Wampold, B., Lichtenberg, J., & Goodyear, R. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*, 218–229.

Trout, J. D. (2008). Seduction without cause: Uncovering explanatory necrophilia. *Trends in Cognitive Sciences,* 12, 281–282.

Vogel, D. (2016, November). Nuclear power and the psychology of evaluating risk. *Skeptical Inquirer, 40*(6), 56–61.

Voss, M. (2012, July). How to spot pseudoneuroscience and biobunk. *APS Observer, 25*, 36–37.

Vyse, S. (2016a). Good news for grouches: Happiness may be overrated. *Skeptical Inquirer, 40*(4), 25–27.

Vyse, S. (2016b, April 28). Syracuse, Apple, and autism pseudoscience. *Skeptical Inquirer: Special Articles.* Retrieved February 13, 2017, from http://www.csicop.org/specialarticles/show/syracuse_apple_and_autism_pseudoscience

Vyse, S. (2017, January). Consensus: Could two hundred scientists be wrong? *Skeptical Inquirer, 41*(1), 29–31.

Waber, R., Shiv, B., Carmon, Z., & Ariely, D. (2008). Commercial features of placebo and therapeutic efficacy. *JAMA, 299*, 1016–1017.

Wade, C., & Tavris, C. (2008). *Psychology* (9th ed.). Upper Saddle River, NJ: Pearson Education.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior.* Hove, England: Erlbaum.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data. *Journal of Personality and Social Psychology, 100*, 426–432.

Walker, J. (2013, October 12). States turn to software to make parole decisions. *Wall Street Journal*, p. A1, A10.

Wang, L. (2009). Money and fame: Vividness effects in the National Basketball Association. *Journal of Behavioral Decision Making, 22*, 20–44.

Wang, Q. (2017, January). Five myths about the role of culture in psychological research. *APS Observer, 30*, 20–24.

12, 281–282.

Tuerkheimer, D. (2010, September 20). Anatomy of a misdiagnosis. *New York Times.* Retrieved November 22, 2010, from http://www.nytimes.com/2010/2009/2021/opinion/2021tuerkheimer.html

University of California (2013, March). Moonstruck. *Berkeley Wellness Letter, 29*(7), 8.

University of California. (2015a, March). Grape expectations. *Berkeley Wellness Letter, 31*(7), 1–2.

University of California. (2015b, Fall Special Issue). The girth of a nation. *Berkeley Wellness Letter, 31*(15), 1–8.

University of California. (2016, February). Mammogram guidelines re-revised. *Berkeley Wellness Letter, 32*(6), 1–2.

Vanderbilt, T. (2008). *Traffic: Why we drive the way we do (and what it says about us)*. New York: Knopf.

Vazire, S., & Gosling, S. D. (2003). Bridging psychology and biology with animal research. *American Psychologist, 58,* 407–408.

Vigen, T. (2015). *Spurious correlations*. New York: Hachette Books.

20–24.

Wargo, E. (2008, October). The many lives of superstition. *APS Observer, 21,* 18–24.

Waterhouse, L. (2006). Multiple intelligences, the Mozart effect, and emotional intelligence: A critical review. *Educational Psychologist, 41*, 207–226.

Watts, D. J. (2011). *Everything is obvious—once you know the answer*. New York: Crown Business.

Wegner, D. A., & Gray, K. (2016). *The mind club*. New York: Viking.

Weir, K. (2015, March). Truth in advertising. *APA Monitor on Psychology, 45,* 36–38.

Weisberg, D. S., Taylor, J. C. V., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgment and Decision Making, 10*, 429–441.

Weisser, C., Renzulli, K., & Leonhardt, M. (2016, June). The 21 most valuable career skills. *Money Magazine*, pp. 42–50.

Welch, H. G., Schwartz, L. M., & Woloshin, S. (2012). *Overdiagnosed: Making people sick in the pursuit of health*. Boston: Beacon Press.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory of mind scale: Longitudinal perspectives. *Child Development, 82,* 780–782.

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect–equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39*, 99–122.

Wessel, D. (2016, October 13). Don't call it a crisis. *Wall Street Journal*, p. 11.

West, S. G. (2009). Alternatives to randomized experiments. *Current Directions in Psychological Science, 18,* 299–304.

Wheelan, C. (2013). *Naked statistics*. New York: Norton.

White, J. (2014, November 24). The hassle of "hands free" car tech. *Wall Street Journal*, p. B1, B4.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322,* 115–117.

Wickens, C., Hollands, J., Banbury, S., & Parasuraman, R. (2012). *Engineering psychology and human performance* (4th ed.). New York: Psychology Press.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences, 112*, 5360–5365.

Willingham, D. T. (2010). *Why don't students like school?* San Francisco: Jossey-Bass.

Willingham, D. T. (2017). *The reading mind*. San Francisco: Jossey-Bass.

Winerman, L. (2014, December). Acing the marshmallow test. *APA Monitor on Psychology, 44,* 28–30.

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*, 515–526.

Wolf, M. (2007). *Proust and the squid.* New York: Harper.

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests, revised-normative update.* San Antonio, TX: Pearson Education.

Youyou, W., Stillwell, D., Schwartz, H., & Kosinski, M. (2017). Birds of a feather do flock together. *Psychological Science, 28*, 276–284.

Zebrowitz, L. A., White, B., & Wieneke, K. (2008). Mere exposure and racial prejudice: Exposure to other-race faces increases liking for strangers of that race. *Social Cognition, 26,* 259–275.

Ziegler, L., & Garfield, J. (2012). Exploring students' intuitive ideas of randomness using an iPod shuffle activity. *Teaching Statistics, 35*, 2–7.

Zimbardo, P. G. (2004). Does psychology make a significant difference in our lives? *American Psychologist, 59,* 339–351.

Zimmer, B. (2010, October 10). Truthiness: The fifth anniversary of Stephen Colbert's introduction of a zeitgeisty word. *New York Times Magazine*, p. 22.

Zuger, A. (2015, August 11). Dr. Fallibility will see you now. *New York Times*, p. D2.

Zvolensky, M. J., Vujanovic, A. A., Bernstein, A., & Leyro, T. (2010). Distress tolerance: Theory, measurement, and relations to psychopathology. *Current Directions in Psychological Science, 19,* 406–410.

Zweig, M. (2008). *Your money and your brain.* New York: Simon & Schuster.

Author Index

Subject Index